

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR
SAMUEL LEMAIRE-PAQUETTE

DEUX NOUVELLES CLASSES DE TESTS D'ADÉQUATION POUR LES
COPULES BASÉS SUR UNE FONCTION CARACTÉRISTIQUE

AVRIL 2019

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

AVANT-PROPOS

En ce qui me concerne, le choix d'un domaine d'étude pour mon parcours universitaire fut un questionnement déchirant. Plusieurs réflexions me poussèrent à renoncer à mon premier choix pour me diriger vers un nouvel univers fascinant : les mathématiques. Malgré une transition difficile et certains doutes en chemin, ce que je me *forçais* à faire était devenu ce que *j'aimais* faire. Parmi l'étendue des champs d'études en mathématiques que j'ai côtoyé, les statistiques présentaient un intérêt tout particulier pour moi. En effet, leur complexité et leur portée d'application dans notre monde ont piqué ma curiosité, me proposant ainsi une perpétuelle quête d'apprentissages et de questionnements. Ce mémoire représente l'ultime accomplissement de mon cheminement scolaire, mais aussi mon épanouissement pour ce merveilleux domaine.

Au terme de ce travail de maîtrise, je tiens à souligner ma reconnaissance envers les professeurs de statistique Mhamed Mesfioui et Nadia Ghazzali, pour m'avoir transmis leur passion et leur expérience, et ce, tout au long de mon parcours universitaire à l'Université du Québec à Trois-Rivières. De plus, je tiens à remercier sincèrement Jean-François Quessy, mon directeur de recherche, pour sa confiance, sa patience et son dévouement. Le partage de son inébranlable expertise en statistique et en recherche m'a permis d'affiner autant ma connaissance que ma rigueur de travail tout en prenant le temps de rire et profiter du moment. Tu as su m'offrir d'innombrables opportunités et je t'en serai éternellement reconnaissant.

Je tiens aussi à remercier ma famille pour leur continuel support. Votre écoute et votre réconfort m'ont soutenu dans les moments difficiles. Votre présence a toujours su me montrer le bon chemin. Finalement, un dernier remerciement tout particulier à ma complice de vie, Valérie, pour son éternel amour. Affronter les nombreux défis de la vie me semble impossible par moments, mais à tes côtés, nous sortons toujours vainqueurs, et ce, tout en grandissant. Merci de m'avoir accompagné dans cette incroyable aventure et de vouloir la poursuivre dans l'inconnu. Après tout, ce n'est que le début.

Table des matières

Avant-propos	ii
Table des matières	iii
Liste des tableaux	vi
Liste des figures	vii
1 Mise en contexte et motivation	1
1.1 Perspective historique	1
1.2 Sur l'importance du choix d'une structure de dépendance	2
1.3 La fonction caractéristique	3
1.4 Utilisation de la fonction caractéristique empirique pour l'inférence . .	6
1.5 Objectifs et structure du mémoire	7
2 Modélisation de la dépendance avec les copules	9
2.1 Les lois multidimensionnelles	9
2.2 Théorème de Sklar et copules	10
2.3 Quelques généralités sur les copules	12
2.3.1 Fonctions de répartition à marges uniformes sur $(0, 1)$	12
2.3.2 La propriété d'invariance d'une copule	13
2.3.3 La copule de survie et la notion de symétrie radiale	13
2.3.4 Les mesures de dépendance	15
2.4 Modèles de copules	17
2.4.1 La copule d'indépendance et les bornes de Fréchet	17

2.4.2	La copule de Farlie–Gumbel–Morgenstern	18
2.4.3	La copule de Plackett	19
2.4.4	La classe des copules Archimédiennes	20
2.4.5	Les copules Normale et Student	21
2.4.6	Les copules Khi-deux et Fisher	23
3	Tests de type fonction caractéristique pour les copules	26
3.1	L'intégrale de probabilité	27
3.2	Pseudo-copies de l'intégrale de probabilité	28
3.3	Test d'adéquation basé sur la fonction de répartition de l'intégrale de probabilité	28
3.3.1	Estimation de K_W	28
3.3.2	Définition de la statistique de test	29
3.3.3	Comportement asymptotique de V_n^K	30
3.3.4	Bootstrap paramétrique pour la p-valeur du test	31
3.4	Tests basés sur la fonction caractéristique de l'intégrale de probabilité .	32
3.4.1	Définition des fonctions caractéristiques	32
3.4.2	Statistiques de test	34
3.4.3	Comportement asymptotique de $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ sous \mathbb{H}_0	40
3.5	Dérivation de formules explicites pour les statistiques de test	42
3.5.1	Formule explicite pour $V_{n,N}^K$	42
3.5.2	Formule explicite pour $S_{n,N}^g$	43
3.5.3	Formule explicite pour $T_{n,N}^\lambda$	44
3.6	Preuve de la Proposition 3.1	44
4	Les nouveaux tests en action	46
4.1	Préliminaires sur les simulations	46
4.1.1	Justification de l'usage d'une méthode Monte–Carlo	46
4.1.2	Statistiques de test étudiées	47
4.1.3	Stratégie pour l'obtention des valeurs critiques	47
4.2	Capacité des tests à conserver leur seuil nominal	50

4.3	Étude sur la puissance des tests	51
4.4	Compléments sur les tests de l'hypothèse d'une copule normale	52
4.5	Illustration sur le jeu de données de Cook & Johnson (1986)	55
4.5.1	Présentation des données	55
4.5.2	Analyse paire par paire	56
4.5.3	Ajustement d'un modèle global à 4 dimensions	59
Conclusion		65
Bibliographie		66
A Éléments complémentaires		73
A.1	Quelques démonstrations	73
A.1.1	Symétrie radiale pour la copule de Frank	73
A.1.2	Dérivation de K_W pour les copules Archimédiennes bidimen- sionnelles	74
A.2	Estimation du paramètre d'une copule	74
A.2.1	Rappel sur les méthodes classiques d'estimation	74
A.2.2	Adaptation aux copules	75
A.3	Détails sur une statistique de test basée sur la copule empirique	77
A.3.1	Estimation non-paramétrique d'une copule	77
A.3.2	Définition de la statistique de test	78
B Tableaux et figures supplémentaires		80

Liste des tableaux

4.1	Tableau-synthèse des modèles de copules acceptés et rejetés pour six paires de variables des données de Cook et Johnson [12].	57
4.2	Résultats du test d'adéquation basé sur $T_{n,N}^{\lambda=3}$ lorsque $N = 250$ pour le vecteur (Co, Cs, Sc, Ti) du jeu de données de Cook et Johnson [12]. . .	59
4.3	Estimations, à l'aide de 1 000 répliqués, des seuils de signification des tests d'adéquation basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$	61
4.4	Estimations, à l'aide de 1 000 répliqués, des pourcentages de rejet de l'hypothèse nulle d'une copule normale pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$	62
4.5	Estimations, à l'aide de 1 000 répliqués, des pourcentages de rejet de l'hypothèse nulle d'une copule de Clayton pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$	63
4.6	Estimations, à l'aide de 1 000 répliqués, des pourcentages de rejet de l'hypothèse nulle d'une copule de Gumbel pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$	64
B.1	Valeurs critiques au seuil $\alpha = 0,05$ pour tester la copule Normale. . . .	80
B.2	P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Co,Cs) et (Co,Sc) du jeu de données de Cook & Johnson. . .	83
B.3	P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Co,Ti) et (Sc,Ti) du jeu de données de Cook & Johnson. . .	84
B.4	P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Cs,Sc) et (Cs,Ti) du jeu de données de Cook & Johnson. . .	85

Table des figures

2.1	Nuages de points (données brutes et paires de rangs standardisés) et histogrammes basés sur 1 000 paires simulées.	11
2.2	Nuage de points de 1 000 paires simulées à partir de la copule de Farlie–Gumbel–Morgenstern lorsque $\tau(C) = 2/9$	18
2.3	Nuages de points de 1 000 paires simulées à partir de la copule de Plackett pour quatre niveaux de dépendance.	19
2.4	Nuages de points de 1 000 paires simulées à partir des copules de Clayton, Gumbel et Frank lorsque $\tau(C) = 1/2$	22
2.5	Nuages de points de 1 000 paires simulées à partir des copules Normale, Khi-deux, Student et Fisher lorsque $\tau(C) = 1/2$	25
3.1	Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ en fonction de $t \in [0, 15]$ pour les copules Normale (ligne 1), Clayton (ligne 2), Gumbel (ligne 3) et Plackett (ligne 4) pour $\tau(C) = 0$ (—), $\tau(C) = .25$ (– –), $\tau(C) = .50$ (– ·) et $\tau(C) = .75$ (· · ·).	35
3.2	Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ en fonction de $t \in [0, 15]$ pour $\tau(C) = .25$ (ligne 1), $\tau(C) = .50$ (ligne 2) et $\tau(C) = .75$ (ligne 3) pour les copules Normale (—), Clayton (– –), Gumbel (– ·) et Plackett (· · ·).	36
3.3	Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W^\lambda(t)$ quand $\lambda = 1$ en fonction de $t \in [0, 2]$ pour les copules Normale (ligne 1), Clayton (ligne 2), Gumbel (ligne 3) et Plackett (ligne 4) pour $\tau(C) = 0$ (—), $\tau(C) = .25$ (– –), $\tau(C) = .50$ (– ·) et $\tau(C) = .75$ (· · ·). . .	37

3.4	Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W^\lambda(t)$ quand $\lambda = 1$ en fonction de $t \in [0, 2]$ pour $\tau(C) = .25$ (ligne 1), $\tau(C) = .50$ (ligne 2) et $\tau(C) = .75$ (ligne 3) pour les copules Normale (—), Clayton (--), Gumbel (-.) et Plackett (···).	38
4.1	Régressions polynomiales en fonction de τ des valeurs critiques pour les statistiques de test $S_{n,N}^{\lambda=1}$ et $T_{n,N}^{\lambda=1}$	50
4.2	Puissance des tests d'adéquation à la copule normale basés sur $S_{n,N}^{\lambda=3}$ (—), $T_{n,N}^{\lambda=3}$ (--) et $V_{n,N}^K$ (-.) en fonction du paramètre de décentralité $a \in [0, 1]$ de la copule Khi-deux bidimensionnelle.	54
4.3	Nuages de points (données brutes et rangs standardisés) et histogrammes pour les variables Co, Cs, Sc et Ti des données de Cook et Johnson [12].	56
4.4	Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ et $\psi_n(t)$ (★★) en fonction de $t \in [0, 15]$ (haut) et de $\psi_W^\lambda(t)$ et $\psi_n^\lambda(t)$ (★★) quand $\lambda = 3$ en fonction de $t \in [0, 1]$ (bas) pour un $\tau(C)$ fixé pour les copules Normale (—) et Gumbel (--).	58
4.5	Nuages de points (données brutes et rangs standardisés) et histogrammes de $n = 655$ observations simulées de la copule de Student à $\nu = 6$ degrés de liberté et de matrice de corrélation $\hat{\Sigma}$ basée sur l'inversion du tau de Kendall à partir des données de Cook et Johnson [12].	60
B.1	Régressions polynomiales des valeurs critiques en fonction de τ pour les statistiques de test $S_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ sous les copules Normale (—), Clayton (--) et Gumbel (-.) lorsque $n = 100$. . .	81
B.2	Régressions polynomiales des valeurs critiques en fonction de τ pour les statistiques de test $S_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ sous les copules Normale (—), Clayton (--) et Gumbel (-.) lorsque $n = 250$. . .	82

Chapitre 1

Mise en contexte et motivation

1.1 Perspective historique

Le concept de copule est assez récent dans l’histoire de la statistique et des probabilités. Encore aujourd’hui, la référence à la « dépendance » est souvent assimilée, à tort, à la très populaire corrélation linéaire de Bravais–Pearson. Certes, la corrélation est une mesure de dépendance relativement simple à calculer et à interpréter. Cependant, strictement parlant, cet indicateur classique est seulement pertinent dans le cas de relations linéaires entre deux ou plusieurs variables aléatoires. De plus, comme il s’agit d’un indice numérique, il ne rend compte que d’un aspect limité de la structure de dépendance entre deux variables aléatoires.

Au contraire, la théorie des copules permet une compréhension beaucoup plus approfondie de la dépendance qui existe entre deux ou plusieurs variables aléatoires. En outre, on verra que cet outil est très flexible pour la construction de modèles multidimensionnels. On voit apparaître les copules pour la première fois dans un article de 1959 du mathématicien américain Abe Sklar [65]. En collaboration avec Berthold

Schweizer, il a découvert le rôle de la fonction de dépendance, *i.e.* la copule, en travaillant sur une version probabiliste de l'espace métrique. Étonnamment, d'autres chercheurs ont pratiquement découvert les copules quelques années avant la célèbre publication de Sklar en travaillant sur des lois multivariées dont les marges sont fixées. On peut citer par exemple les travaux de Höffding [41], Fréchet [27], Dall'Aglia [15] et Féron [24].

Curieusement, en dépit de l'importance de la découverte de Sklar, il y a eu, au départ, peu d'engouements pour les copules. En fait, il a fallu attendre quelque deux décennies avant que les copules ne constituent un sujet de recherche de pointe en statistique et probabilités. Un premier pas a été franchi avec les travaux de Kimeldorf et Sampson publiés entre 1975 et 1987 (par exemple [48]), de même qu'avec ceux de Deheuvels (voir [16], [17] et [18]). Une autre impulsion est venue des travaux du statisticien québécois Christian Genest et de ses nombreux collaborateurs qui ont étudié de façon intensive les propriétés théoriques des copules et mis au point une panoplie de méthodologies statistiques pour faire de l'inférence. Plusieurs domaines ont bénéficié de ces travaux, notamment la finance, l'actuariat, l'hydrologie et la climatologie.

1.2 Sur l'importance du choix d'une structure de dépendance

Pour bien comprendre un phénomène qui met en scène deux ou plusieurs variables aléatoires, il est crucial de choisir une structure de dépendance qui colle bien aux données observées. À ce titre, les tests d'adéquation pour les copules sont des outils indispensables. L'objectif de ces tests est de vérifier formellement si la copule sous-jacente à une population multidimensionnelle appartient ou non à une certaine famille paramétrique de copules, qui peut s'exprimer sous la forme $\{C_\theta; \theta \in \Theta\}$, où $\Theta \subseteq \mathbb{R}^p$ est l'espace des paramètres.

Dans ce contexte, les hypothèses nulle et alternative à confronter s'expriment par

$$\mathbb{H}_0 : C \in \{C_\theta; \theta \in \Theta\} \quad \text{et} \quad \mathbb{H}_1 : C \notin \{C_\theta; \theta \in \Theta\}.$$

Les premiers auteurs à s'intéresser à tester les hypothèses \mathbb{H}_0 et \mathbb{H}_1 précédentes sont Genest et Rivest [35]. Une formalisation de leurs tests fut proposée par Quessy [59] et Genest *et al.* [33] via l'utilisation de processus empiriques. Plusieurs autres tests ont été proposés au fil des ans, notamment Scaillet [62], Berg [7], Mesfioui *et al.* [55], Omelka *et al.* [57], Genest *et al.* [29] et Genest *et al.* [31]. Pour le lecteur intéressé, un recensement de quelques-uns de ces tests, accompagné d'une étude de simulations étendue, se retrouvent dans Genest *et al.* [37].

1.3 La fonction caractéristique

Ce mémoire propose de développer de nouveaux tests d'adéquation pour les copules autour de la notion de fonction caractéristique qui peut être associée à la loi d'une variable aléatoire ou d'un vecteur de variables aléatoires. Pour la définir de façon générale, prenons un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ dans \mathbb{R}^d . On sait que la loi de \mathbf{X} peut se caractériser avec sa fonction de répartition, c'est-à-dire par

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Sur la base d'un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ dans \mathbb{R}^d , un estimateur non-paramétrique naturel pour $F_{\mathbf{X}}$ est la fonction de répartition empirique

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(\mathbf{X}_j \leq \mathbf{x}).$$

Outre le fait que $F_n(\mathbf{x})$ est un estimateur sans biais pour $F_{\mathbf{X}}(\mathbf{x})$, on sait aussi que F_n est uniformément convergente pour $F_{\mathbf{X}}$ en probabilité, au sens où pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sup_{\mathbf{x} \in \mathbb{R}^d} |F_n(\mathbf{x}) - F_{\mathbf{X}}(\mathbf{x})| < \epsilon \right) = 1.$$

Il existe une façon alternative de caractériser la loi de \mathbf{X} , c'est-à-dire en considérant sa fonction caractéristique. En définissant l'unité imaginaire $i = \sqrt{-1}$, la fonction caractéristique de \mathbf{X} est définie pour $\mathbf{t} = (t_1, \dots, t_d) \in \mathbb{R}^d$ par

$$\psi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E} \left(e^{i\mathbf{t}\mathbf{X}^\top} \right). \quad (1.1)$$

Dans l'équation (1.1) de même que pour le reste du mémoire, \mathbf{a}^\top dénote la transposée du vecteur $\mathbf{a} = (a_1, \dots, a_d)$. À l'instar de $F_{\mathbf{X}}$, la fonction $\psi_{\mathbf{X}}$ caractérise complètement la loi de \mathbf{X} . En effet, on a $\psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{Y}}(\mathbf{t})$ pour tout $\mathbf{t} \in \mathbb{R}^d$ si et seulement si les vecteurs aléatoires \mathbf{X} et \mathbf{Y} sont égaux en loi. Si \mathbf{X} possède une densité $f_{\mathbf{X}}$, le Théorème d'inversion stipule que celle-ci se déduit de sa fonction caractéristique via

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\mathbf{t}\mathbf{x}^\top} \psi_{\mathbf{X}}(\mathbf{t}) d\mathbf{t}.$$

L'estimateur non-paramétrique naturel de $\psi_{\mathbf{X}}$ est la fonction caractéristique empirique telle que proposée par Feuerverger et Mureika [25] et étudiée par Csörgő [13] et Csörgő [14], c'est-à-dire

$$\psi_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n e^{i\mathbf{t}\mathbf{X}_j^\top}. \quad (1.2)$$

Tout comme F_n pour $F_{\mathbf{X}}$, on montre aisément que $\psi_n(\mathbf{t})$ est sans biais pour $\psi_{\mathbf{X}}(\mathbf{t})$. De plus, ψ_n converge presque sûrement vers $\psi_{\mathbf{X}}$ au sens où pour tout $B \in (0, \infty)$,

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \sup_{|\mathbf{t}| \leq B} |\psi_n(\mathbf{t}) - \psi_{\mathbf{X}}(\mathbf{t})| = 0 \right) = 1.$$

Ce résultat se retrouve, entre autres, dans Csörgő [13] pour $d = 1$ et dans Csörgő [14] pour son extension quand $d > 1$.

La fonction caractéristique $\psi_{\mathbf{X}}$ et sa version empirique ψ_n sont des fonctions à valeurs dans l'espace des nombres complexes; elles possèdent ainsi une partie réelle et une partie imaginaire. En partant de l'identité $e^{ix} = \cos x + i \sin x$, on peut donc écrire que $\psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{X}}^{\text{Re}}(\mathbf{t}) + i \psi_{\mathbf{X}}^{\text{Im}}(\mathbf{t})$, où

$$\psi_{\mathbf{X}}^{\text{Re}}(\mathbf{t}) = \mathbb{E} \{ \cos(\mathbf{t} \mathbf{X}^\top) \} \quad \text{et} \quad \psi_{\mathbf{X}}^{\text{Im}}(\mathbf{t}) = \mathbb{E} \{ \sin(\mathbf{t} \mathbf{X}^\top) \}.$$

De manière similaire, on a $\psi_n(\mathbf{t}) = \psi_n^{\text{Re}}(\mathbf{t}) + i \psi_n^{\text{Im}}(\mathbf{t})$, où

$$\psi_n^{\text{Re}}(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n \cos(\mathbf{t} \mathbf{X}_j^\top) \quad \text{et} \quad \psi_n^{\text{Im}}(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n \sin(\mathbf{t} \mathbf{X}_j^\top).$$

On peut déduire les moments d'une variable aléatoire $X \in \mathbb{R}$ à partir de sa fonction caractéristique ψ_X . En effet, on montre facilement que

$$\lim_{t \rightarrow 0} \psi_X^{(p)}(t) = i^p \mathbb{E}(X^p).$$

Exemple 1.1. *Un cas simple survient lorsque X est de loi uniforme sur $(0, 1)$. On montre alors facilement que $\psi_X(t) = (e^{it} - 1)/(it)$. Par conséquent,*

$$\psi_X^{\text{Re}}(t) = \frac{\sin(t)}{t} \quad \text{et} \quad \psi_X^{\text{Im}}(t) = \frac{1 - \cos(t)}{t}.$$

Exemple 1.2. *La fonction caractéristique d'une variable aléatoire X de loi Normale de moyenne μ et de variance σ^2 est $\psi_X(t) = \exp(it\mu - \sigma^2 t^2/2)$. On a alors*

$$\psi_X^{\text{Re}}(t) = \cos(t\mu) e^{-\sigma^2 t^2/2} \quad \text{et} \quad \psi_X^{\text{Im}}(t) = \sin(t\mu) e^{-\sigma^2 t^2/2}.$$

Dans le cas particulier de la loi standard ($\mu = 0$, $\sigma^2 = 1$), on a

$$\psi_X^{\text{Re}}(t) = e^{-t^2/2} \quad \text{et} \quad \psi_X^{\text{Im}}(t) = 0.$$

Pour déduire les moments de la normale, on note que

$$\psi_X^{(1)}(t) = (i\mu - \sigma^2 t) \psi_X(t) \quad \text{et} \quad \psi_X^{(2)}(t) = -\sigma^2 t + (i\mu - \sigma^2) \psi_X^{(1)}(t).$$

Par conséquent, comme $\psi_X(0) = 1$, on a $\psi_X^{(1)}(0) = i\mu$ et $\psi_X^{(2)}(0) = (i\mu)^2 = -\sigma^2 - \mu^2$.

Il en découle que

$$E(X) = \psi_X^{(1)}(0)/i = \mu \quad \text{et} \quad E(X^2) = \psi_X^{(2)}(0)/i^2 = \sigma^2 + \mu^2.$$

1.4 Utilisation de la fonction caractéristique empirique pour l'inférence

Au même titre qu'il y a deux approches pour représenter le comportement d'un vecteur aléatoire, c'est-à-dire via $F_{\mathbf{X}}$ ou $\psi_{\mathbf{X}}$, il y a deux catégories générales de statistiques pour tester des hypothèses relatives au comportement de \mathbf{X} : celles basées sur F_n , et celles basées sur ψ_n . Dans la suite, on énumérera plusieurs travaux qui se sont servis de la fonction caractéristique pour élaborer des méthodologies statistiques.

Tout d'abord, des tests pour l'hypothèse de symétrie de \mathbf{X} autour d'un point $\mathbf{a} \in \mathbb{R}^d$ ont été développés par Henze *et al.* [40] ; ils exploitent le fait que sous l'hypothèse nulle $\mathbb{H}_0 : \mathbf{X} - \mathbf{a} \stackrel{d}{=} \mathbf{a} - \mathbf{X}$, on a $\psi_{\mathbf{Y}}(\mathbf{t}) = \psi_{\mathbf{Y}}(-\mathbf{t})$, où $\mathbf{Y} = \mathbf{X} - \mathbf{a}$. Spécifiquement, leur test est basé sur le fait que $\psi_{\mathbf{Y}}(\mathbf{t}) - \psi_{\mathbf{Y}}(-\mathbf{t}) = 2E\{\sin(\mathbf{t}\mathbf{Y}^\top)\}$. Pour tester l'égalité en loi de deux vecteurs aléatoires \mathbf{X} et \mathbf{Y} , Alba Fernández *et al.* [1] ont utilisé la forme de l'hypothèse basée sur les fonctions caractéristiques $\psi_{\mathbf{X}}(t) = E(e^{it\mathbf{X}^\top})$ et $\psi_{\mathbf{Y}}(t) = E(e^{it\mathbf{Y}^\top})$ respectives à \mathbf{X} et à \mathbf{Y} , à savoir $\mathbb{H}_0 : \psi_{\mathbf{X}}(\mathbf{t}) = \psi_{\mathbf{Y}}(\mathbf{t})$. Enfin, Meintanis et Iliopoulos [53] ont élaboré des tests pour l'indépendance entre les composantes d'un vecteur $\mathbf{X} = (X_1, \dots, X_d)$.

Plusieurs tests d'adéquation basés sur la fonction caractéristique ont également été développés. Par exemple, Epps et Pulley [21] ont proposé des tests de normalité dans le cas unidimensionnel, alors que Baringhaus et Henze [6] élargissent leur applicabilité pour la normalité multidimensionnelle. Des tests d'adéquation pour des mélanges de lois normales sont offerts par Klar et Meintanis [49], alors que les travaux de Kim et Genton [47] se concentrent sur les mélanges de distributions normales *asymétrisées* et Epps [20] sur les lois unidimensionnelles de type lieu-échelle. Des tests pour la loi de Cauchy sont proposés par Gürtler et Henze [39] et par Matsui et Takemura [51], alors que Fragiadakis et Meintanis [26] se concentrent sur la distribution de Laplace multidimensionnelle. Une approche unificatrice permettant de tester une famille arbitraire de modèles est élaborée par Jiménez-Gamero *et al.* [44]. Une version plus rapide du point de vue calculatoire a été décrite par Jiménez-Gamero et Kim [43].

1.5 Objectifs et structure du mémoire

Plusieurs des tests décrits à la Section 1.4 sont plus performants que des versions similaires basées sur la fonction de répartition. Malgré ces succès, peu de chercheurs se sont attardés jusqu'ici à l'utilisation de fonctions caractéristiques dans un contexte de modélisation par les copules. Ce travail de maîtrise vise à combler en partie cette lacune en développant de nouveaux tests d'adéquation à une copule basés sur des fonctions caractéristiques définies à partir de la variable aléatoire *Transformation intégrale de probabilité* (TIP) multidimensionnelle. Formellement, pour un vecteur aléatoire $\mathbf{X} \in \mathbb{R}^d$ de fonction de répartition $F_{\mathbf{X}}$, la TIP est la variable aléatoire $W = F_{\mathbf{X}}(\mathbf{X})$. Les tests de Genest et Rivest [35] et de Genest *et al.* [33] sont d'ailleurs basés sur W . Toutefois, leur méthodologie concernant les tests est construite autour de la *fonction de répartition* de W , à savoir

$$F_W(w) = \mathbb{P}(W \leq w).$$

Ce mémoire propose donc de travailler plutôt avec la *fonction caractéristique* de W , c'est-à-dire $\psi_W(t) = E(e^{itW})$. Cette idée n'a jamais été exploitée auparavant dans la littérature sur les copules. On espère ainsi, à l'instar des tests décrits à la Section 1.4, que des approches utilisant ψ_W plutôt que F_W seront plus performantes. En outre, une version pondérée de ψ_W sera proposée, dans la lignée d'une suggestion récente faite par Meintanis *et al.* [52].

Le mémoire est structuré comme suit. Au Chapitre 2, le concept de copule est introduit et plusieurs propriétés sont ensuite décrites ; de nombreux modèles sont ensuite détaillés. Le lecteur qui est familier avec les copules pourra passer directement au chapitre suivant. Les chapitres 3 et 4 constituent la contribution originale de ce travail de maîtrise. Au Chapitre 3, on définit deux types de fonctions caractéristiques associées à la variable aléatoire de la transformation intégrale de probabilité multidimensionnelle ; ces fonctions sont ensuite utilisées pour élaborer de nouvelles méthodologies pour l'adéquation à une famille de copules. Le Chapitre 4 étudie les propriétés échantillonnelles des nouveaux tests ; en particulier, des comparaisons sont effectuées avec deux tests d'adéquation très populaires. Une illustration sur de vraies données multidimensionnelles est également fournie. Pour permettre une lecture plus fluide, les technicalités mathématiques, de même que certains tableaux et certaines figures, sont consignés dans une annexe.

Chapitre 2

Modélisation de la dépendance avec les copules

2.1 Les lois multidimensionnelles

La copule d'un vecteur de variables aléatoires est intimement liée à sa fonction de répartition multidimensionnelle. Afin de formaliser le tout, soit un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ dans \mathbb{R}^d dont la fonction de répartition est définie par

$$F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Pour chaque $\ell \in \{1, \dots, d\}$, les fonctions de répartition marginales F_1, \dots, F_d sont données par $F_{\ell}(x) = \mathbb{P}(X_{\ell} \leq x)$. Ces dernières se déduisent de $F_{\mathbf{X}}$ par

$$F_{\ell}(x_{\ell}) = \lim_{x_j \rightarrow \infty: j \neq \ell} F_{\mathbf{X}}(\mathbf{x}).$$

Quand la loi de \mathbf{X} est continue, le Théorème de Radon–Nikodym (voir par exemple Casella et Berger [10] ou Shao [63]) assure l'existence d'une fonction non-négative

$f_{\mathbf{X}} : \mathbb{R}^d \rightarrow \mathbb{R}$ appelée densité de probabilité conjointe de \mathbf{X} telle que

$$\int_{\mathbb{R}^d} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x} = 1 \quad \text{et} \quad \mathbb{P}(\mathbf{X} \in \mathcal{A}) = \int_{\mathcal{A}} f_{\mathbf{X}}(\mathbf{x}) \, d\mathbf{x}.$$

En particulier, on retrouve $F_{\mathbf{X}}(\mathbf{x})$ en posant $\mathcal{A} = (-\infty, x_1] \times \cdots \times (-\infty, x_d]$.

2.2 Théorème de Sklar et copules

Le résultat fondamental de Sklar [65] permet de comprendre le rôle des copules dans la relation entre une fonction de répartition multidimensionnelle et chacune de ses fonctions de répartition marginales. Ce résultat est formellement énoncé ci-dessous.

Théorème 2.1 (Théorème de Sklar). *Soit une fonction de répartition d -dimensionnelle $F_{\mathbf{X}}$ dont les marges F_1, \dots, F_d sont continues. Alors il existe une unique fonction $C : [0, 1]^d \rightarrow [0, 1]$ telle que pour tout $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, on a la représentation*

$$F_{\mathbf{X}}(\mathbf{x}) = C \{F_1(x_1), \dots, F_d(x_d)\}. \quad (2.1)$$

La fonction C s'appelle la copule de $F_{\mathbf{X}}$.

Le Théorème 2.1 permet de constater que la loi de \mathbf{X} est composée des comportements individuels de X_1, \dots, X_d à travers les lois marginales F_1, \dots, F_d , et de la dépendance entre les d composantes de \mathbf{X} , représentée par C . Autrement dit, si l'on s'intéresse aux liens d'interdépendance entre les composantes d'un vecteur aléatoire, alors on peut se concentrer uniquement sur sa copule. La Figure 2.1 représente une paire aléatoire (X_1, X_2) simulée 1 000 fois pour laquelle la loi de X_1 est Gamma (histogramme en haut à gauche) et la loi de X_2 est Normale (histogramme en bas à droite). En haut à droite, on retrouve le nuage de points basé sur les données brutes, alors qu'en bas à gauche, on découvre la structure de dépendance, c'est-à-dire la forme de la copule.

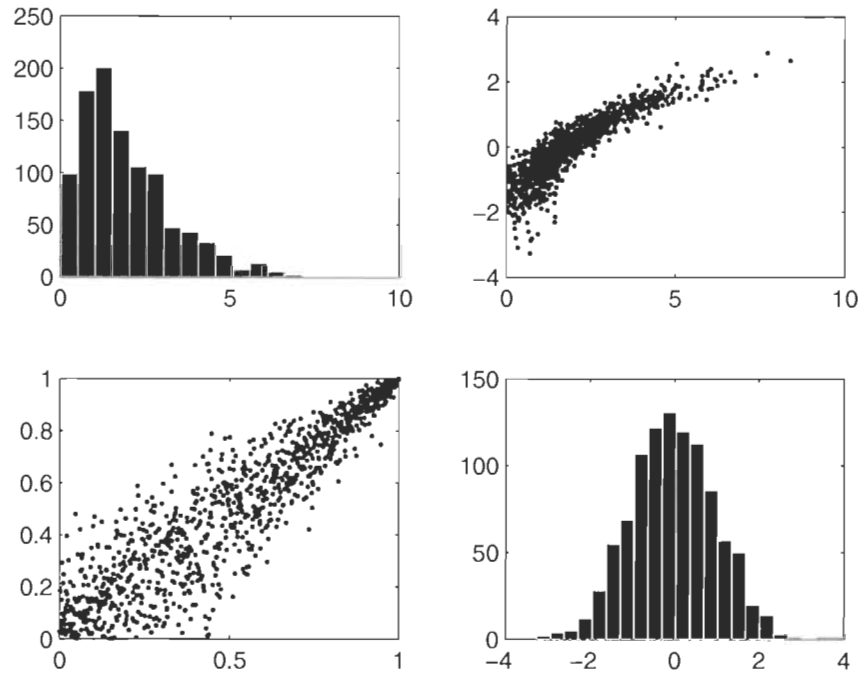


FIGURE 2.1 – Nuages de points (données brutes et paires de rangs standardisés) et histogrammes basés sur 1 000 paires simulées.

Une autre utilité du Théorème de Sklar est la possibilité d'extraire l'unique copule associée à une fonction de répartition $F_{\mathbf{X}}$. En effet, en posant $u_\ell = F_\ell(x_\ell)$ pour chaque $\ell \in \{1, \dots, d\}$ dans l'équation (2.1), on déduit que la copule de $F_{\mathbf{X}}$ est donnée pour $\mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d$ par

$$C(\mathbf{u}) = F_{\mathbf{X}} \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}, \quad (2.2)$$

où F^{-1} représente l'inverse de la fonction de répartition F . Un exemple plutôt simple pour illustrer l'extraction d'une copule à partir d'une fonction de répartition est présenté dans la suite.

Exemple 2.1. Soit la fonction de répartition bivariée $F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}$ définie sur $[0, \infty)^2$. D'abord, ses fonctions de répartition marginales sont données par

$$F_1(x_1) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) = \frac{1}{1 + e^{-x_1}} \quad \text{et} \quad F_2(x_2) = \lim_{x_1 \rightarrow \infty} F(x_1, x_2) = \frac{1}{1 + e^{-x_2}}.$$

En notant que $F_1^{-1}(u) = F_2^{-1}(u) = -\ln(-1 + 1/u)$, une application de (2.2) amène

$$C(u_1, u_2) = F \left\{ -\ln \left(-1 + \frac{1}{u_1} \right), -\ln \left(-1 + \frac{1}{u_2} \right) \right\} = \frac{u_1 u_2}{u_1 + u_2 - u_1 u_2}.$$

2.3 Quelques généralités sur les copules

2.3.1 Fonctions de répartition à marges uniformes sur $(0, 1)$

Il est intéressant de noter que la copule C associée à un vecteur aléatoire \mathbf{X} de loi $F_{\mathbf{X}}$ et de marges continues F_1, \dots, F_d correspond à la loi conjointe de $\mathbf{U} = (U_1, \dots, U_d)$, où $U_\ell = F_\ell(X_\ell)$ pour chaque $\ell \in \{1, \dots, d\}$. En effet, un calcul direct donne

$$\begin{aligned} \mathbb{P}(U_1 \leq u_1, \dots, U_d \leq u_d) &= \mathbb{P}\{F_1(X_1) \leq u_1, \dots, F_d(X_d) \leq u_d\} \\ &= \mathbb{P}\{X_1 \leq F_1^{-1}(u_1), \dots, X_d \leq F_d^{-1}(u_d)\} \\ &= F_{\mathbf{X}}\{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\} \\ &= C\{F_1 \circ F_1^{-1}(u_1), \dots, F_d \circ F_d^{-1}(u_d)\} \\ &= C(\mathbf{u}), \end{aligned}$$

où $F \circ F^{-1}(u)$ représente la composition de fonctions entre la fonction de répartition et son inverse. On voit donc que C est elle-même une fonction de répartition d -dimensionnelle, puisqu'elle correspond à la loi conjointe de \mathbf{U} . En outre, ses marges sont uniformes sur $(0, 1)$, car d'un résultat bien connu en statistique mathématique, la loi de $U_\ell = F_\ell(X_\ell)$ est justement l'uniforme sur $(0, 1)$. On pourrait ainsi définir la fonction caractéristique associable à une copule C à partir de \mathbf{U} , c'est-à-dire $\psi_C(\mathbf{t}) = E(e^{i\mathbf{t}\mathbf{U}^\top})$. La fonction ψ_C a été employée par Bahraoui et Quessy [4] pour tester la symétrie radiale d'une copule, par Bahraoui *et al.* [2] pour tester l'adéquation, et par Bahraoui *et al.* [3] pour tester l'échangeabilité bidimensionnelle d'une copule.

2.3.2 La propriété d'invariance d'une copule

L'invariance sous des transformations monotones croissantes est une propriété fondamentale des copules. Elle stipule que si $\beta_1, \dots, \beta_d : \mathbb{R} \rightarrow \mathbb{R}$ sont des fonctions monotones et croissantes, alors la copule de $\tilde{\mathbf{X}} = (\beta_1(X_1), \dots, \beta_d(X_d))$ est la même que celle de $\mathbf{X} = (X_1, \dots, X_d)$. Pour établir cette propriété, on note d'abord que

$$\begin{aligned} F_{\tilde{\mathbf{X}}}(\mathbf{x}) &= \mathbb{P}\{\beta_1(X_1) \leq x_1, \dots, \beta_d(X_d) \leq x_d\} \\ &= \mathbb{P}\{X_1 \leq \beta_1^{-1}(x_1), \dots, X_d \leq \beta_d^{-1}(x_d)\} \\ &= F_{\mathbf{X}}\{\beta_1^{-1}(x_1), \dots, \beta_d^{-1}(x_d)\}. \end{aligned}$$

En notant C la copule de \mathbf{X} , on a $F_{\mathbf{X}}(\mathbf{x}) = C\{F_1(x_1), \dots, F_d(x_d)\}$ et on peut écrire

$$F_{\tilde{\mathbf{X}}}(\mathbf{x}) = C\{F_1 \circ \beta_1^{-1}(x_1), \dots, F_d \circ \beta_d^{-1}(x_d)\}.$$

Puisque les marges de $\tilde{\mathbf{X}}$ sont justement $F_1 \circ \beta_1^{-1}(x_1), \dots, F_d \circ \beta_d^{-1}(x_d)$, on déduit de l'expression précédente et du Théorème de Sklar que C est également la copule de $\tilde{\mathbf{X}}$.

2.3.3 La copule de survie et la notion de symétrie radiale

La fonction de survie est un concept fondamental en statistique possédant de nombreuses applications. Dans le cas d'une variable aléatoire continue $X \in \mathbb{R}$ de fonction de répartition $F_X(x) = \mathbb{P}(X \leq x)$, sa fonction de survie est définie par

$$\bar{F}_X(t) = \mathbb{P}(X > t) = 1 - F_X(t).$$

Un exemple classique concerne la durée de vie X d'un individu ; dans ce cas, $\bar{F}(t)$ est la probabilité qu'il survive au-delà d'un certain temps $t > 0$. Dans le cas d'un vecteur

aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ dans \mathbb{R}^d , la fonction de survie conjointe est définie par

$$\bar{F}_{\mathbf{X}}(t_1, \dots, t_d) = \mathbb{P}(X_1 > t_1, \dots, X_d > t_d).$$

De même que l'on peut identifier une fonction de dépendance, c'est-à-dire une copule, liant les lois marginales d'une fonction de répartition multidimensionnelle, on peut aussi montrer qu'il existe une fonction $\hat{C} : [0, 1]^d \rightarrow [0, 1]$ telle que

$$\mathbb{P}(X_1 > x_1, \dots, X_d > x_d) = \hat{C} \{ \bar{F}_1(x_1), \dots, \bar{F}_d(x_d) \},$$

où $\bar{F}_1, \dots, \bar{F}_d$ sont les fonctions de survie marginales. Il s'agit en quelque sorte d'une version *survie* du Théorème de Sklar [65]. La fonction \hat{C} s'appelle justement la copule de survie. L'expression de \hat{C} en termes de C est compliquée dans le cas général $d > 2$. Toutefois, dans le cas bivarié, on a $\hat{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$.

Par ailleurs, la symétrie autour de $\mathbf{a} = (a_1, \dots, a_d) \in \mathbb{R}^d$ d'un vecteur aléatoire \mathbf{X} est liée à sa copule C , de même qu'à sa copule de survie \hat{C} . D'abord, dire que \mathbf{X} est symétrique autour de $\mathbf{a} \in \mathbb{R}^d$ signifie que $\mathbf{X} - \mathbf{a}$ et $\mathbf{a} - \mathbf{X}$ possèdent la même loi, c'est-à-dire $\mathbb{P}(\mathbf{X} - \mathbf{a} \leq \mathbf{x}) = \mathbb{P}(\mathbf{a} - \mathbf{X} \leq \mathbf{x})$ pour tout $\mathbf{x} \in \mathbb{R}^d$. Si $F_{\mathbf{X}}$ est la fonction de répartition conjointe de \mathbf{X} et $\bar{F}_{\mathbf{X}}$ sa fonction de survie, on obtient $F_{\mathbf{X}}(\mathbf{x} + \mathbf{a}) = \bar{F}_{\mathbf{X}}(\mathbf{a} - \mathbf{x})$. En invoquant le Théorème de Sklar et sa version survie, on a alors

$$C \{ F_1(x_1 + a_1), \dots, F_d(x_d + a_d) \} = \hat{C} \{ \bar{F}_1(a_1 - x_1), \dots, \bar{F}_d(a_d - x_d) \}.$$

Donc, \mathbf{X} sera symétrique autour de \mathbf{a} si et seulement si $C = \hat{C}$ et

$$F_{\ell}(x_{\ell} + a_{\ell}) = \bar{F}_{\ell}(a_{\ell} - x_{\ell})$$

pour chaque $\ell \in \{1, \dots, d\}$, c'est-à-dire que X_{ℓ} est symétrique autour de a_{ℓ} . La symétrie de $\mathbf{X} = (X_1, \dots, X_d)$ est donc équivalente aux symétries marginales de X_1, \dots, X_d et à $C = \hat{C}$. Quand $C = \hat{C}$, la copule C possède la propriété de symétrie radiale.

2.3.4 Les mesures de dépendance

La corrélation de Bravais–Pearson est probablement la mesure la plus connue lorsqu’il est question de mesurer la dépendance entre deux variables aléatoires. Plus spécifiquement, la corrélation entre deux variables aléatoires X_1 et X_2 est définie par

$$\rho = \frac{\mathbb{E} \{ (X_1 - \mathbb{E}(X_1)) (X_2 - \mathbb{E}(X_2)) \}}{\sqrt{\text{var}(X_1) \text{var}(X_2)}}.$$

Toutefois, cette mesure souffre de quelques lacunes. Notamment, elle est convenable seulement lorsque le lien entre X_1 et X_2 est linéaire. De plus, la valeur de ρ dépend des lois marginales de X_1 et de X_2 .

À la lumière du Théorème de Sklar, une propriété souhaitable pour une mesure de dépendance entre deux variables aléatoires X_1 et X_2 serait qu’elle dépende uniquement de la copule sous-jacente à leur loi conjointe, donc qu’elle ne soit pas influencée par les lois marginales. À ce titre, deux mesures de dépendance très populaires se démarquent, à savoir le tau de Kendall et le rho de Spearman ; elles sont décrites dans ce qui suit.

Le tau de Kendall tel qu’initialement formulé par Kendall [46] est défini par la différence des probabilités de concordance et de discordance entre deux couples indépendants (X_1, X_2) et (X_1^*, X_2^*) de même loi $F_{(X_1, X_2)}$, à savoir

$$\tau = \mathbb{P} \{ (X_1 - X_1^*)(X_2 - X_2^*) > 0 \} - \mathbb{P} \{ (X_1 - X_1^*)(X_2 - X_2^*) < 0 \}.$$

Dans le cas où la loi $F_{(X_1, X_2)}$ est continue, on peut montrer que τ ne dépend que de sa copule C via la formule

$$\tau(C) = 4 \int_0^1 \int_0^1 C(u_1, u_2) \, dC(u_1, u_2) - 1. \quad (2.3)$$

Le rho de Spearman ρ_S d'un couple (X_1, X_2) est défini comme la corrélation entre $U_1 = F_1(X_1)$ et $U_2 = F_2(X_2)$. Sachant que dans le cas continu, la loi conjointe de (U_1, U_2) est la copule C de (X_1, X_2) , on déduit alors que

$$\rho_S(C) = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3. \quad (2.4)$$

Dans un autre ordre d'idée, les notions de mesure de dépendance codale s'intéressent à la propension qu'ont deux variables X_1 et X_2 à prendre simultanément des valeurs extrêmes. L'indice codal supérieur mesure la probabilité que X_1 et X_2 soient simultanément grandes via la formule

$$\lambda_U = \lim_{t \rightarrow 1^-} \mathbb{P} \{ X_1 > F_2^{-1}(t) | X_1 > F_1^{-1}(t) \}.$$

De manière semblable, l'indice codal inférieur mesure la probabilité que X_1 et X_2 soient simultanément petites, c'est-à-dire

$$\lambda_L = \lim_{t \rightarrow 0^+} \mathbb{P} \{ X_2 \leq F_2^{-1}(t) | X_1 \leq F_1^{-1}(t) \}.$$

Dans le cas de variables continues, λ_U et λ_L ne dépendent que de la copule C de (X_1, X_2) . En effet, on peut montrer que

$$\lambda_U(C) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \quad \text{et} \quad \lambda_L(C) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}.$$

Puisque $\lambda_U(C)$ et $\lambda_L(C)$ sont des probabilités, on a $\lambda_U(C), \lambda_L(C) \in [0, 1]$. Lorsque $\lambda_U(C) > 0$, on dit que les variables X_1 et X_2 sont asymptotiquement dépendantes supérieurement ; elles sont dites asymptotiquement dépendantes inférieurement quand $\lambda_L(C) > 0$. Dans les autres cas, on dit qu'elles sont asymptotiquement indépendantes.

2.4 Modèles de copules

2.4.1 La copule d'indépendance et les bornes de Fréchet

La structure de dépendance la plus simple apparaît lorsque les composantes d'un vecteur $\mathbf{X} = (X_1, \dots, X_d)$ sont indépendantes. Dans ce cas, on sait que $F_{\mathbf{X}}(\mathbf{x}) = F_1(x_1) \times \dots \times F_d(x_d)$. Par la formule (2.2), on déduit que la copule associée à \mathbf{X} est

$$\Pi(\mathbf{u}) = F_{\mathbf{X}} \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\} = u_1 \times \dots \times u_d.$$

La fonction Π s'appelle la copule d'indépendance multidimensionnelle. Dans le cas bivarié, $\Pi(u_1, u_2) = u_1 u_2$ et il est facile de montrer via les formules (2.3) et (2.4) que le tau de Kendall et le rho de Spearman sont nuls, c'est-à-dire que $\tau(\Pi) = \rho_S(\Pi) = 0$.

Alors que Π correspond à l'indépendance entre des variables aléatoires, la borne de Fréchet supérieure est associée à la dépendance positive parfaite. Dans ce cas, les variables X_1, \dots, X_d sont liées de telle sorte que pour chaque $\ell \in \{2, \dots, d\}$, on a $X_\ell = \beta_\ell(X_1)$ presque sûrement pour β_2, \dots, β_d monotones croissantes. On a donc

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \mathbb{P} \{X_1 \leq x_1, \beta_2(X_1) \leq x_2, \dots, \beta_d(X_1) \leq x_d\} \\ &= \mathbb{P} \{X_1 \leq \min(x_1, \beta_2^{-1}(x_2), \dots, \beta_d^{-1}(x_d))\}. \end{aligned}$$

Puisque les marges de $F_{\mathbf{X}}$ sont respectivement $x_1, \beta_2^{-1}(x_2), \dots, \beta_d^{-1}(x_d)$, une application de la formule (2.2) montre que la copule dans ce cas est $M(\mathbf{u}) = \min(u_1, \dots, u_d)$. Il n'y a cependant pas de définition claire de la notion de dépendance négative parfaite dans le cas à $d > 2$ variables. Dans le cas $d = 2$, toutefois, on dit que les variables X_1 et X_2 sont parfaitement dépendantes négativement si $X_2 = \tilde{\beta}(X_1)$ presque sûrement, où $\tilde{\beta}$ est monotone décroissante. Dans ce cas, la copule est $W(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$; cette structure de dépendance s'appelle la borne inférieure de Fréchet.

2.4.2 La copule de Farlie–Gumbel–Morgenstern

La copule de Farlie–Gumbel–Morgenstern à $d = 2$ variables est définie par $C_\theta(u_1, u_2) = u_1u_2 + \theta u_1(1 - u_1)u_2(1 - u_2)$, où $\theta \in [-1, 1]$. Il existe des versions à $d > 2$ dimensions de cette copule, mais elles sont limitées du point de vue de leur utilité. À partir des formules (2.3) et (2.4), on montre que $\tau(C_\theta) = 2\theta/9$ et $\rho_S(C_\theta) = \theta/3$. Ainsi, on constate que la plage des dépendances possibles dans ce modèle est limitée, car

$$-\frac{2}{9} \leq \tau(C_\theta) \leq \frac{2}{9} \quad \text{et} \quad -\frac{1}{3} \leq \rho_S(C_\theta) \leq \frac{1}{3}.$$

La copule de Farlie–Gumbel–Morgenstern possède la propriété de symétrie radiale. En effet, par un calcul direct, on montre facilement que

$$\widehat{C}_\theta(u_1, u_2) = u_1 + u_2 - 1 + C_\theta(1 - u_1, 1 - u_2) = u_1u_2 + \theta u_1u_2(1 - u_1)(1 - u_2) = C_\theta(u_1, u_2).$$

La Figure 2.2 montre un nuage de points de 1 000 paires simulées à partir de cette copule lorsque $\theta = 1$, donc quand $\tau(C_\theta) = 2/9$. On peut constater sur cette figure que la force de la dépendance est assez faible. D'autre part, on voit que la dépendance est symétrique du point de vue des queues inférieure et supérieure ; il s'agit en fait de la manifestation de la propriété de symétrie radiale.

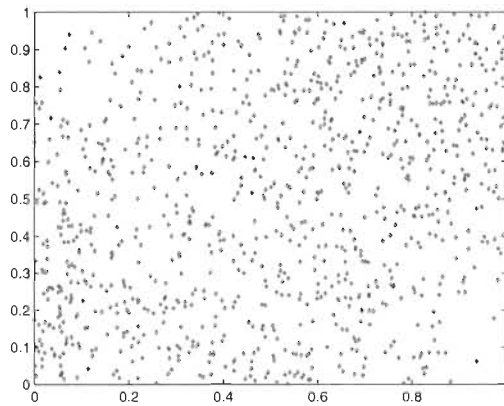


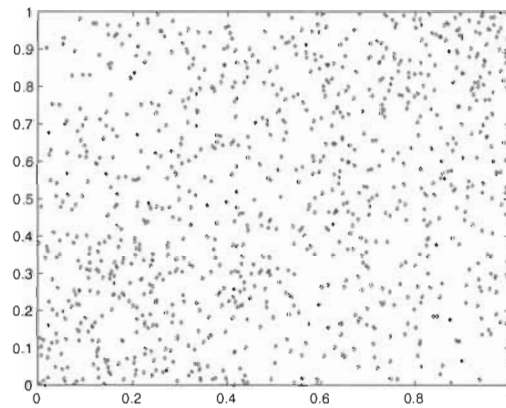
FIGURE 2.2 – Nuage de points de 1 000 paires simulées à partir de la copule de Farlie–Gumbel–Morgenstern lorsque $\tau(C) = 2/9$.

2.4.3 La copule de Plackett

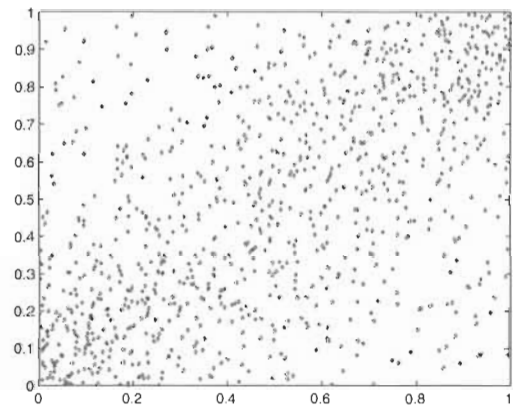
La copule de Plackett bvariée est extraite d'une loi bidimensionnelle telle que décrite par Plackett [58]. Son expression est donnée pour $\theta > 0$ par

$$C_\theta(u_1, u_2) = \frac{Q_\theta(u_1, u_2) - \sqrt{Q_\theta^2(u_1, u_2) - 4\theta(\theta - 1)u_1u_2}}{2(\theta - 1)},$$

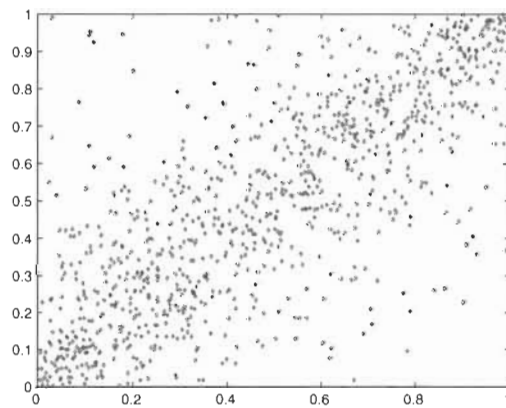
où $Q_\theta(u_1, u_2) = 1 + (\theta - 1)(u_1 + u_2)$. On retrouve la copule d'indépendance comme limite de C_θ lorsque $\theta \rightarrow 1$. La Figure 2.3 montre les nuages de points obtenus à partir de paires simulées de cette copule pour quatre niveaux de dépendance.



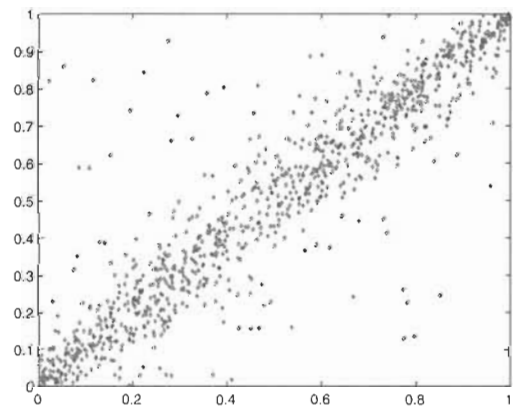
(a) $\tau(C) = 1/5$



(b) $\tau(C) = 2/5$



(c) $\tau(C) = 3/5$



(d) $\tau(C) = 4/5$

FIGURE 2.3 – Nuages de points de 1 000 paires simulées à partir de la copule de Plackett pour quatre niveaux de dépendance.

2.4.4 La classe des copules Archimédiennes

On dit qu'une copule multidimensionnelle C appartient à la famille Archimédienne si elle peut s'écrire sous la forme

$$C_\phi(\mathbf{u}) = \phi^{-1} \left\{ \sum_{j=1}^d \phi(u_j) \right\},$$

où $\phi : [0, 1] \rightarrow [0, \infty)$ s'appelle le générateur. Cette classe générale de copules a été introduite par Genest et MacKay [32]. À noter que ϕ doit satisfaire $\phi(1) = 0$ et pour $g(t) = \phi^{-1}(t)$,

$$(-1)^\ell \frac{d^\ell}{dt^\ell} g(t) \geq 0.$$

En dimension deux, $C_\phi(u_1, u_2) = \phi^{-1} \{ \phi(u_1) + \phi(u_2) \}$ et on peut montrer que

$$\tau(C_\phi) = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} dt.$$

On peut aussi établir que

$$\lambda_U(C_\phi) = 2 - \lim_{x \rightarrow 0^+} \frac{1 - \phi^{-1}(2x)}{1 - \phi^{-1}(x)} \quad \text{et} \quad \lambda_L(C_\phi) = \lim_{x \rightarrow \infty} \frac{\phi^{-1}(2x)}{\phi^{-1}(x)}.$$

Quelques copules Archimédiennes sont décrites dans la suite.

Exemple 2.2 (Copule de Clayton). *La copule de Clayton s'extrait d'un modèle multidimensionnel initialement étudié par Clayton [11]. Le générateur de cette copule est défini pour $\theta > 0$ par $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$. On montre alors que cette copule s'écrit*

$$C_\theta(\mathbf{u}) = \left(\sum_{j=1}^d u_j^{-\theta} - d + 1 \right)^{-1/\theta}.$$

On peut montrer que le tau de Kendall de la copule de Clayton bidimensionnelle est $\tau(C_\theta) = \theta/(\theta + 2)$. La copule de Clayton ne possède pas de dépendance codale supérieure, à savoir que $\lambda_U(C_\theta) = 0$, mais $\lambda_L(C_\theta) = 2^{-1/\theta}$.

Exemple 2.3 (Copule de Gumbel). *La copule de Gumbel telle qu'extraite du modèle bidimensionnel logistique considéré par Gumbel [38] et Hougaard [42] est générée par $\phi_\theta(t) = |\ln t|^{1/(1-\theta)}$, où $\theta \in [0, 1]$. La copule de Gumbel est donc de la forme*

$$C_\theta(\mathbf{u}) = \exp \left\{ - \left(\sum_{j=1}^d |\ln u_j|^{1/(1-\theta)} \right)^{1-\theta} \right\}.$$

Elle possède l'unique particularité d'appartenir à la fois aux familles Archimédienne et à valeurs extrêmes; ce résultat a été démontré par Genest et Rivest [34].

Exemple 2.4 (Copule de Frank). *La copule de Frank a été initialement étudiée par Genest [30]; son générateur est défini par $\phi_\theta(t) = \ln(1 - e^{-t}) - \ln(1 - e^{-\theta t})$. Ainsi,*

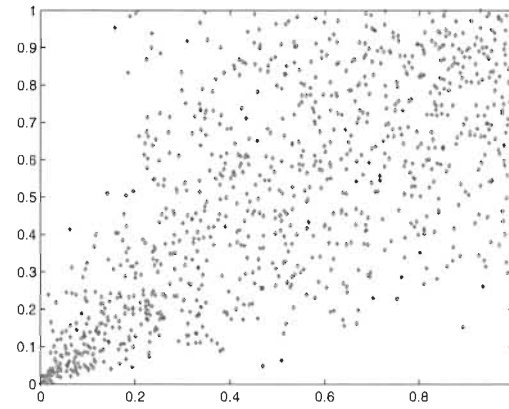
$$C_\theta(\mathbf{u}) = -\frac{1}{\theta} \ln \left\{ 1 - \frac{1}{1 - e^{-\theta}} \prod_{j=1}^d (1 - e^{-\theta u_j}) \right\}.$$

Il est montré dans l'annexe à la Section A.1.1 que la copule de Frank bidimensionnelle possède la propriété de symétrie radiale, à savoir que $C_\theta(u_1, u_2) = \widehat{C}_\theta(u_1, u_2)$. En fait, on pourrait même démontrer que la copule de Frank est la seule, parmi la classe Archimédienne, à posséder cette caractéristique. À noter enfin que cette copule ne possède pas de dépendance codale, c'est-à-dire que $\lambda_L(C_\theta) = \lambda_U(C_\theta) = 0$.

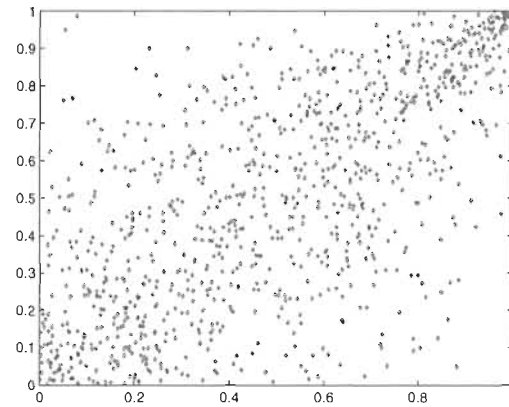
Pour clore cette section, la Figure 2.4 montre les nuages de points de paires simulées à partir des copules de Clayton, Gumbel et Frank lorsque $\tau(C) = 1/2$.

2.4.5 Les copules Normale et Student

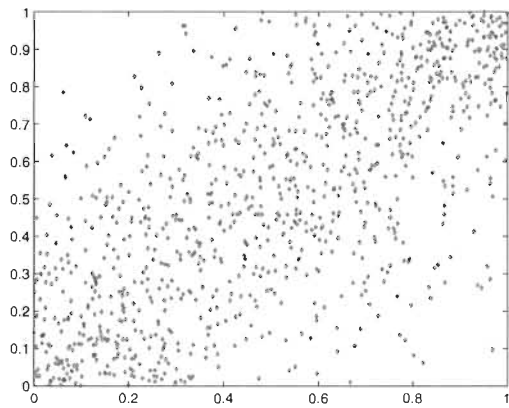
La copule Normale est la structure de dépendance extraite de la loi Normale multidimensionnelle; elle possède une formulation implicite. Pour la décrire, soit Φ_Σ , la fonction de répartition de la Normale à d dimensions de moyennes nulles, de variances unitaires, et de matrice de corrélation $\Sigma \in \mathbb{R}^{d \times d}$. En posant ϕ_Σ la densité associée à



(a) Copule de Clayton



(b) Copule de Gumbel



(c) Copule de Frank

FIGURE 2.4 – Nuages de points de 1 000 paires simulées à partir des copules de Clayton, Gumbel et Frank lorsque $\tau(C) = 1/2$.

Φ_Σ , on a la représentation implicite

$$\Phi_\Sigma(\mathbf{x}) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} \phi_\Sigma(\mathbf{s}) \, ds_1 \cdots ds_d.$$

Puisque toutes les marges de Φ_Σ correspondent à la fonction de répartition Φ de la loi $\mathcal{N}(0, 1)$, on déduit que la copule Normale s'exprime par

$$C_\Sigma(\mathbf{u}) = \Phi_\Sigma \{ \Phi^{-1}(u_1), \dots, \Phi^{-1}(u_d) \} = \int_{-\infty}^{\Phi^{-1}(u_d)} \cdots \int_{-\infty}^{\Phi^{-1}(u_1)} \phi_\Sigma(\mathbf{s}) \, ds_1 \cdots ds_d.$$

La copule de Student à ν degrés de liberté s'obtient de la même façon. Ainsi, en posant que $F_{\Sigma, \nu}$ est la fonction de répartition de la loi de Student d -dimensionnelle et que F_ν est la fonction de répartition de la loi de Student unidimensionnelle à ν degrés de liberté, on déduit la représentation implicite de la copule de Student, à savoir

$$C_{\Sigma, \nu}(\mathbf{u}) = F_{\Sigma, \nu} \{ F_\nu^{-1}(u_1), \dots, F_\nu^{-1}(u_d) \}.$$

Les Figures 2.5a et 2.5c montrent respectivement des nuages de points de données simulées selon ces deux modèles de copules.

2.4.6 Les copules Khi-deux et Fisher

Soit un vecteur aléatoire $\mathbf{Z} = (Z_1, \dots, Z_d)$ de loi Normale de moyennes nulles, de variances unitaires, et de matrice de corrélation Σ . La copule Khi-deux est définie par la structure de dépendance extraite du vecteur aléatoire $((Z_1 + a)^2, \dots, (Z_d + a)^2)$, où $a \in \mathbb{R}$ est un paramètre de décentralité. Ce modèle de copules a d'abord été proposé par Bârdossy [8], alors que ses propriétés ont été étudiées par Quessy *et al.* [61]. En outre, on retrouve la copule Normale lorsque $a \rightarrow \infty$. Il est démontré par Quessy *et al.*

[61] que la copule Khi-deux multidimensionnelle peut s'exprimer sous la forme

$$C_{\Sigma,a}^x(\mathbf{u}) = \sum_{\epsilon \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j \right) \Phi_{\Sigma} \{h_a(\epsilon_1 u_1), \dots, h_a(\epsilon_d u_d)\},$$

où pour $G_{a_j}(x) = \Phi(\sqrt{x} - a_j) + \Phi(\sqrt{x} + a_j) - 1$, on a

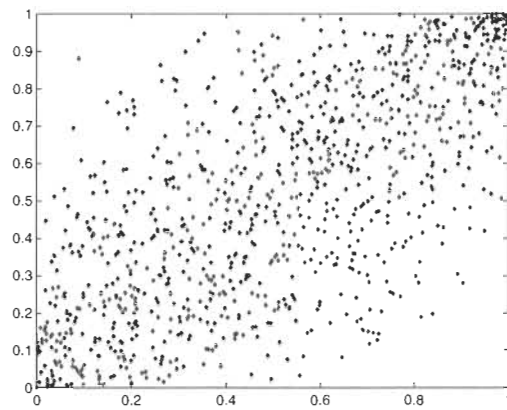
$$h_a(u) = \text{sign}(u) \sqrt{G_a^{-1}(|u|)} - a.$$

L'avantage de cette copule par rapport à la structure de dépendance normale repose sur le fait qu'elle peut accommoder de l'asymétrie radiale.

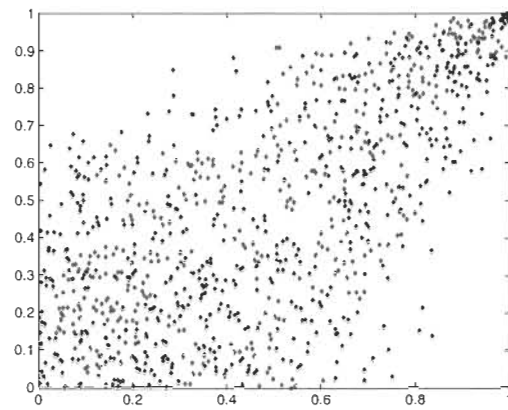
La copule de Fisher utilise une construction similaire à celle qui mène à la copule Khi-deux centrée, *i.e.* lorsque $a = 0$. Ainsi, telle que proposée par Favre *et al.* [23], la copule de Fisher à ν degrés de liberté et de matrice de corrélation Σ est définie comme la structure de dépendance du vecteur (X_1^2, \dots, X_d^2) dans le cas où (X_1, \dots, X_d) est distribué selon la loi de Student à ν degrés de liberté. Cette copule peut s'écrire

$$C_{\Sigma,\nu}^F(\mathbf{u}) = \sum_{\epsilon \in \{-1,1\}^d} \left(\prod_{j=1}^d \epsilon_j \right) C_{\Sigma,\nu}^t \left(\frac{1 + \epsilon_1 u_1}{2}, \dots, \frac{1 + \epsilon_d u_d}{2} \right).$$

Outre le fait que la copule de Fisher possède la propriété d'asymétrie radiale, elle permet de la dépendance de queue supérieure, contrairement à la copule Khi-deux. À noter qu'à la limite lorsque $\nu \rightarrow \infty$, on retrouve la copule Khi-deux centrée ($a = 0$). Les Figures 2.5b et 2.5d montrent des nuages de points de données simulées selon les copules Khi-deux centrée et Fisher à $\nu = 5$ degrés de liberté.



(a) Copule Normale



(b) Copule Khi-deux centrée

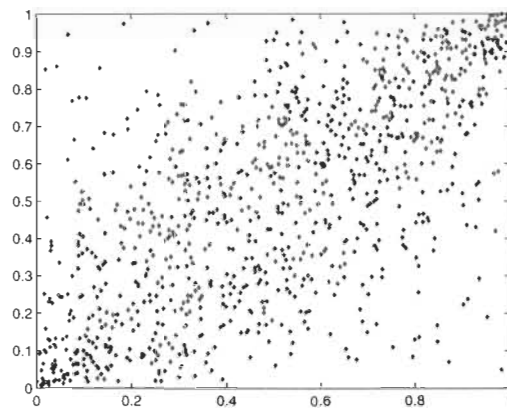
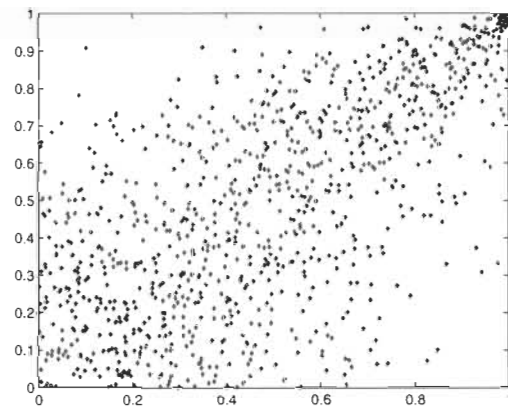
(c) Copule de Student ($\nu = 5$)(d) Copule de Fisher ($\nu = 5$)

FIGURE 2.5 – Nuages de points de 1 000 paires simulées à partir des copules Normale, Khi-deux, Student et Fisher lorsque $\tau(C) = 1/2$.

Chapitre 3

Tests de type fonction caractéristique pour les copules

Le chapitre précédent a permis de nous familiariser avec le concept de copules et ses propriétés associées, mais aussi avec les différentes structures que celles-ci peuvent prendre. Pour la suite des choses, il devient logique d'aborder la notion de tests d'adéquation. Soit un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. d'une loi $F_{\mathbf{X}}$ dont les marges sont continues. Cette loi possède donc une unique copule C , en vertu du Théorème de Sklar. Le but d'un test d'adéquation pour une copule est de vérifier s'il est plausible que la copule C de $F_{\mathbf{X}}$ appartienne à la famille $\{C_{\theta}; \theta \in \Theta\}$, où $\theta = (\theta_1, \dots, \theta_p)$ est un paramètre inconnu. Formellement, l'hypothèse nulle et alternative à confronter sont

$$\mathbb{H}_0 : C \in \{C_{\theta}; \theta \in \Theta\} \quad \text{et} \quad \mathbb{H}_1 : C \notin \{C_{\theta}; \theta \in \Theta\}.$$

L'objectif de ce chapitre consiste à définir de nouveaux tests pour choisir entre \mathbb{H}_0 et \mathbb{H}_1 . On verra que ces tests sont, d'une certaine manière, des versions naturelles à la méthodologie de Genest et Rivest [35] et Genest *et al.* [33] basée sur l'intégrale de probabilité multidimensionnelle. On introduira d'abord formellement la notion d'inté-

grale de probabilité pour ainsi établir son lien avec les copules. Les détails concernant la méthodologie de Genest *et al.* [33] seront ensuite donnés. La table sera alors mise afin d'introduire les nouveaux tests d'adéquation pour les copules.

3.1 L'intégrale de probabilité

Telle que définie au Chapitre 1, la variable aléatoire *Transformation intégrale de probabilité* (TIP) d'un vecteur aléatoire $\mathbf{X} \in \mathbb{R}^d$ de fonction de répartition $F_{\mathbf{X}}$ est la variable aléatoire $W = F_{\mathbf{X}}(\mathbf{X})$. En fait, dans le cas où $F_{\mathbf{X}}$ est continue, la variable W ne dépend que de la copule C de $F_{\mathbf{X}}$. En effet, du Théorème de Sklar,

$$W = F_{\mathbf{X}}(\mathbf{X}) = C\{F_1(X_1), \dots, F_d(X_d)\} = C(U_1, \dots, U_d),$$

où $\mathbf{U} = (U_1, \dots, U_d) \sim C$. On peut donc définir la TIP d'un vecteur \mathbf{X} par $W = C(\mathbf{U})$. La fonction de répartition de W est notée $K_W(w) = \mathbb{P}(W \leq w)$, où $w \in [0, 1]$. Dans le cas bivarié, Genest et Rivest [35] ont montré que

$$K_W(w) = w + \int_w^1 C^{(01)}\{C_v^{-1}(t), v\} dv, \quad (3.1)$$

où

$$C^{(01)}(u_1, u_2) = \frac{\partial}{\partial u_2} C(u_1, u_2) \quad \text{et} \quad C_v^{-1}(t) = \inf\{u : C(u, v) \geq t\}.$$

En particulier, on a dans le cas d'une copule Archimédienne de générateur ϕ que

$$K_W(w) = w - \frac{\phi(w)}{\phi'(w)}. \quad (3.2)$$

Pour davantage de détails, cette formule est démontrée à la Section A.1.2 en annexe. Par exemple, dans le cas de la copule de Clayton présentée à l'Exemple 2.2, on a

$$K_W(w) = w + \frac{w(1-w^\theta)}{\theta}.$$

3.2 Pseudo-copies de l'intégrale de probabilité

À partir d'un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. de loi $F_{\mathbf{X}}$, on pourrait former l'échantillon W_1, \dots, W_n i.i.d. K_W , où pour chaque $j \in \{1, \dots, n\}$, $W_j = F_{\mathbf{X}}(\mathbf{X}_j)$ serait une copie de l'intégrale de probabilité $W = F_{\mathbf{X}}(\mathbf{X})$. Cependant, cette façon de procéder est inapplicable, puisque $F_{\mathbf{X}}$ est inconnue. Pour y remédier, on se basera sur l'estimation naturelle de $F_{\mathbf{X}}$ fournie par la fonction de répartition empirique

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(\mathbf{X}_j \leq \mathbf{x}).$$

On définit alors l'échantillon de *pseudo-observations* $\widehat{W}_1, \dots, \widehat{W}_n$, où pour chaque $j \in \{1, \dots, n\}$, $\widehat{W}_j = F_n(\mathbf{X}_j)$. On a donc

$$\widehat{W}_j = \frac{1}{n} \sum_{\ell=1}^n \mathbf{1}(\mathbf{X}_{\ell} \leq \mathbf{X}_j).$$

3.3 Test d'adéquation basé sur la fonction de répartition de l'intégrale de probabilité

3.3.1 Estimation de K_W

L'approche préconisée par Genest et Rivest [35] et Genest *et al.* [33] est basée sur la fonction de répartition K_W de l'intégrale de probabilité. Une estimation de K_W basée sur les *pseudo-observations* $\widehat{W}_1, \dots, \widehat{W}_n$ est alors donnée par

$$K_n(w) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(\widehat{W}_j \leq w). \quad (3.3)$$

Les propriétés asymptotiques de K_n en tant qu'estimateur de K_W ont été étudiées minutieusement par Barbe *et al.* [5]. Ces auteurs montrent que le processus empirique

$$\mathbb{K}_n(w) = \sqrt{n} \{K_n(w) - K_W(w)\}$$

converge en loi vers un processus gaussien dont la fonction de covariance dépend de la copule de \mathbf{X} . Ce résultat est valide sous l'Hypothèse I suivante.

Hypothèse I. La densité k_W de K_W est continue sur $(0, 1]$ et pour tout $\epsilon > 0$,

$$k_W(w) = o \left\{ w^{-1/2} \log^{-1/2-\epsilon} \left(\frac{1}{w} \right) \right\}$$

lorsque $w \rightarrow 0$. De plus, il existe une version de la fonction de répartition conditionnelle de \mathbf{X} sachant que $W = w$ telle que pour toutes fonctions continues $f : \mathbb{R}^d \rightarrow \mathbb{R}$, l'application $w \mapsto \mu(w, f) = k_W(w) \mathbb{E} \{f(\mathbf{X}) | W = w\}$ est continue sur $(0, 1]$, alors que $\mu(1, f) = k_W(1) f(1, \dots, 1)$.

3.3.2 Définition de la statistique de test

L'idée de Genest *et al.* [33] est de mesurer une distance entre la fonction de répartition empirique de l'intégrale de probabilité, c'est-à-dire K_n , et sa version K_θ sous l'hypothèse nulle que $C \in \{C_\theta; \theta \in \Theta\}$. Ainsi, une des statistiques de test considérées par Genest *et al.* [33] est basée sur la fonctionnelle de Cramer-von Mises, à savoir

$$V_n^K = n \int_0^1 \{K_n(w) - K_{\hat{\theta}_n}(w)\}^2 dw. \quad (3.4)$$

Dans cette expression, $\hat{\theta}_n$ est un estimateur convergent pour θ . Des détails concernant les estimateurs possibles pour θ sont donnés dans l'Annexe A.2.

Le calcul de V_n^K s'avère parfois difficile, voire même impossible, lorsque K_θ n'admet pas de forme explicite. Pour pallier ce problème, Genest et Rémillard [36] proposent d'utiliser une version de V_n^K dans laquelle $K_{\hat{\theta}_n}$ est estimée. Précisément, ces auteurs suggèrent de se baser sur la statistique

$$V_{n,N}^K = n \int_0^1 \{K_n(w) - K_N(w)\}^2 dw, \quad (3.5)$$

où K_N est la fonction de répartition empirique calculée à partir d'observations simulées $\mathbf{X}_1^*, \dots, \mathbf{X}_N^*$ i.i.d. $C_{\hat{\theta}_n}$, desquelles on déduit les pseudo-observations $\widehat{W}_1^*, \dots, \widehat{W}_N^*$. La statistique $V_{n,N}^K$ a l'avantage de se calculer facilement. En effet, il est montré à la Section 3.5 que

$$\begin{aligned} V_{n,N}^K &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \left\{ 1 - \max(\widehat{W}_j, \widehat{W}_k) \right\} \\ &\quad - \frac{2}{N} \sum_{j=1}^n \sum_{k=1}^N \left\{ 1 - \max(\widehat{W}_j, \widehat{W}_k^*) \right\} \\ &\quad + \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \left\{ 1 - \max(\widehat{W}_j^*, \widehat{W}_k^*) \right\}. \end{aligned} \quad (3.6)$$

3.3.3 Comportement asymptotique de V_n^K

En définissant le processus empirique $\widetilde{\mathbb{K}}_n(w)$, on obtient une expression intéressante pour V_n^K . En d'autres termes, on trouve respectivement

$$\widetilde{\mathbb{K}}_n(w) = \sqrt{n} \{K_n(w) - K_{\hat{\theta}_n}(w)\} \quad \text{et} \quad V_n^K = \int_0^1 \left\{ \widetilde{\mathbb{K}}_n(w) \right\}^2 dw.$$

Sous l'Hypothèse I déjà décrite et sous l'Hypothèse II ci-dessous, Genest *et al.* [33] ont démontré que sous \mathbb{H}_0 , le processus $\widetilde{\mathbb{K}}_n$ converge vers la limite gaussienne

$$\widetilde{\mathbb{K}}(w) = \mathbb{K}(t) - \dot{K}_\theta(w)^\top A,$$

où \mathbb{K} est le processus limite de \mathbb{K}_n et A est la limite en loi de $A_n = \sqrt{n}(\hat{\theta}_n - \theta) \in \mathbb{R}^p$. À noter que dans l'expression précédente, la fonction $\dot{K}_\theta(w)$ est définie par la dérivée partielle de $K_\theta(w)$ par rapport à θ , à savoir

$$\dot{K}_\theta(w) = \frac{\partial}{\partial \theta} K_\theta(w).$$

Du Théorème des applications continues, on déduit alors directement que sous \mathbb{H}_0 , V_n^K converge en loi vers la variable aléatoire

$$\mathbb{V}^K = \int_0^1 \left\{ \tilde{\mathbb{K}}(w) \right\}^2 dw.$$

Hypothèse II. Le processus (\mathbb{K}_n, A_n) converge en loi dans l'espace $D[0, 1] \times \mathbb{R}^p$ vers une limite gaussienne et centrée (\mathbb{K}, A) . De plus, pour tout $\theta \in \mathcal{O} \subseteq \mathbb{R}^p$, la fonction $\dot{K}_\theta(w)$ existe et est continue pour tout $w \in [0, 1]$, alors que pour $\epsilon \rightarrow 0$,

$$\sup_{\|\theta^* - \theta\| < \epsilon} \sup_{w \in [0, 1]} \left| \dot{K}_{\theta^*}(w) - \dot{K}_\theta(w) \right| \rightarrow 0.$$

À noter que sous quelques conditions de régularités additionnelles, Genest et Rémillard [36] montrent que le résultat de convergence de V_n^K vers \mathbb{V}^K est aussi valide pour sa version estimée $V_{n,N}^K$. On sait donc que $V_{n,N}^K$ converge en loi sous \mathbb{H}_0 vers une limite non-dégénérée. Cependant, la forme de celle-ci est loin de posséder une forme explicite, ce qui rend difficile la déduction directe des valeurs critiques.

3.3.4 Bootstrap paramétrique pour la p-valeur du test

La règle de décision concernant les hypothèses d'adéquation \mathbb{H}_0 et \mathbb{H}_1 consiste à rejeter \mathbb{H}_0 si la statistique de test $V_{n,N}^K$ prend une valeur *trop élevée*. Il y a cependant des obstacles à l'utilisation d'une telle règle de décision. D'une part, la loi limite \mathbb{V}^K de $V_{n,N}^K$ est intraitable, et d'autre part cette dernière dépend de toute façon du paramètre

θ , qui est inconnu. Pour ces raisons, une version de la méthode de ré-échantillonnage du *bootstrap paramétrique* telle qu'initialement proposée par Stute *et al.* [66] sera employée. Dans le cas de $V_{n,N}^K$, sa validité a été démontrée par Genest et Rémillard [36]. Les étapes de sa mise en œuvre sont décrites ci-dessous.

- (1) **Calcul de la statistique.** À partir des observations $\mathbf{X}_1, \dots, \mathbf{X}_n$, on calcule l'estimateur $\hat{\theta}_n$ et la statistique de test $V_{n,N}^K$;
- (2) **Obtention de répliques de $V_{n,N}^K$.** Pour chaque $m \in \{1, \dots, M\}$, on génère un échantillon $\mathbf{X}_1^{(m)}, \dots, \mathbf{X}_n^{(m)}$ i.i.d. $C_{\hat{\theta}_n}$; on calcule ensuite l'estimateur $\hat{\theta}_n^{(m)}$ et la statistique de test $V_{n,N}^{(m)}$;
- (3) **Estimation de la p-valeur du test et prise de décision.** À partir des répliques $V_{n,N}^{(1)}, \dots, V_{n,N}^{(M)}$, on estime la p-valeur du test avec

$$\widehat{PV} = \frac{1}{M} \sum_{m=1}^M \mathbf{1} \left(V_{n,N}^{(m)} > V_{n,N}^K \right).$$

Enfin, pour un seuil α désiré, on rejette \mathbb{H}_0 si $\widehat{PV} < \alpha$.

3.4 Tests basés sur la fonction caractéristique de l'intégrale de probabilité

3.4.1 Définition des fonctions caractéristiques

La version de la fonction caractéristique de W telle que présentée à l'Équation (1.1) est définie pour $t \in \mathbb{R}$ par

$$\psi_W(t) = \mathbb{E} \left(e^{itW} \right).$$

À l'instar de F_W , la fonction ψ_W caractérise le comportement de W . Bien entendu, ψ_W possède une partie réelle et une partie imaginaire, données respectivement par

$$\psi_W^{\text{Re}}(t) = \mathbb{E} \{ \cos(tW) \} = \int_0^1 \cos(tw) k_W(w) dw$$

et

$$\psi_W^{\text{Im}}(t) = \mathbb{E} \{ \sin(tW) \} = \int_0^1 \sin(tw) k_W(w) dw.$$

En général, il semble difficile, voire impossible, d'obtenir des formules explicites pour ψ_W^{Re} et ψ_W^{Im} . Une exception survient lorsque $C = M$, où $M(u_1, u_2) = \min(u_1, u_2)$ est la borne de Fréchet supérieure. En effet, si $(U_1, U_2) \sim M$, alors $U_1 = U_2$ presque sûrement et donc $W = M(U_1, U_2) = \min(U_1, U_1) = U_1$ est uniforme sur $(0, 1)$. Ainsi, ψ_W est la même que la fonction caractéristique de la loi Uniforme décrite à l'Exemple 1.1.

Pour contourner le fait que des versions explicites de ψ_W^{Re} et ψ_W^{Im} sont ardues à obtenir pour une copule C quelconque, on utilisera les approximations

$$\psi_W^{\text{Re}}(t) \approx \frac{1}{B} \sum_{\ell=1}^B \cos(tW_\ell) \quad \text{et} \quad \psi_W^{\text{Im}}(t) \approx \frac{1}{B} \sum_{\ell=1}^B \sin(tW_\ell),$$

où pour $(U_{11}, U_{12}), \dots, (U_{B1}, U_{B2})$ i.i.d. C , on a posé $W_\ell = C(U_{\ell 1}, U_{\ell 2})$ pour chaque $\ell \in \{1, \dots, B\}$. Dans ce qui suit, on a utilisé $B = 5\,000$. La Figure 3.1 présente les courbes de ψ_W^{Re} et ψ_W^{Im} séparément pour les copules Normale, Clayton, Gumbel et Plackett en faisant varier le niveau de dépendance; on constate alors l'influence du tau de Kendall sur le comportement de ces courbes. La Figure 3.2 montre plutôt les différences entre les fonctions caractéristiques pour ces quatre modèles, cette fois en fixant la valeur du tau de Kendall. D'une certaine manière, chaque copule possède sa propre *signature*; les différences entre les modèles sont plutôt notables lorsque $|t| \geq 5$.

On considère également une version pondérée de ψ_W , à savoir

$$\psi_W^\lambda(t) = \mathbb{E} \left\{ (W(1-W))^{\lambda|t|} e^{itW} \right\}.$$

Cette fonction caractéristique est inspirée d'une proposition récente faite par Meintanis *et al.* [52] et explorée subséquemment par Meintanis et Ushakov [54]. On verra plus loin l'avantage de travailler avec une telle version pondérée. Similairement à ψ_W^{Re} et ψ_W^{Im} , on propose d'approximer la partie réelle et imaginaire de ψ_W^λ avec

$$\psi_W^{\lambda, \text{Re}}(t) \approx \frac{1}{B} \sum_{\ell=1}^B (W_\ell(1 - W_\ell))^{\lambda|t|} \cos(tW_\ell)$$

et

$$\psi_W^{\lambda, \text{Im}}(t) \approx \frac{1}{B} \sum_{\ell=1}^B (W_\ell(1 - W_\ell))^{\lambda|t|} \sin(tW_\ell).$$

Les figures 3.3 et 3.4 sont des analogues aux figures 3.1 et 3.2, mais pour ψ_W^λ lorsque $\lambda = 1$. On remarque très peu de variation pour les parties réelles de ψ_W^λ , que ce soit en faisant varier $\tau(C)$ ou en comparant les copules entre elles. Ainsi, les différences sont plutôt notables au niveau de la partie imaginaire de ψ_W^λ .

3.4.2 Statistiques de test

À partir des pseudo-observations $\widehat{W}_1, \dots, \widehat{W}_n$, des estimateurs naturels de ψ_W et ψ_W^λ sont respectivement

$$\psi_n(t) = \frac{1}{n} \sum_{j=1}^n e^{it\widehat{W}_j} \quad \text{et} \quad \psi_n^\lambda(t) = \frac{1}{n} \sum_{j=1}^n \left(\widehat{W}_j(1 - \widehat{W}_j) \right)^{\lambda|t|} e^{it\widehat{W}_j}.$$

Soit maintenant ψ_θ et ψ_θ^λ , la fonction caractéristique et la version pondérée de la fonction caractéristique de la copule C_θ présente dans la famille $\{C_\theta; \theta \in \Theta\}$ qui est testée sous \mathbb{H}_0 . En posant $|z|^2 = a^2 + b^2$ le module au carré du nombre complexe $a + ib$, les statistiques de test proposées pour confronter \mathbb{H}_0 à \mathbb{H}_1 sont

$$S_n^g = n \int_{\mathbb{R}} |\psi_n(t) - \psi_{\widehat{\theta}}(t)|^2 g(t) dt \quad \text{et} \quad T_n^\lambda = n \int_{\mathbb{R}} |\psi_n^\lambda(t) - \psi_{\widehat{\theta}_n}^\lambda(t)|^2 dt.$$

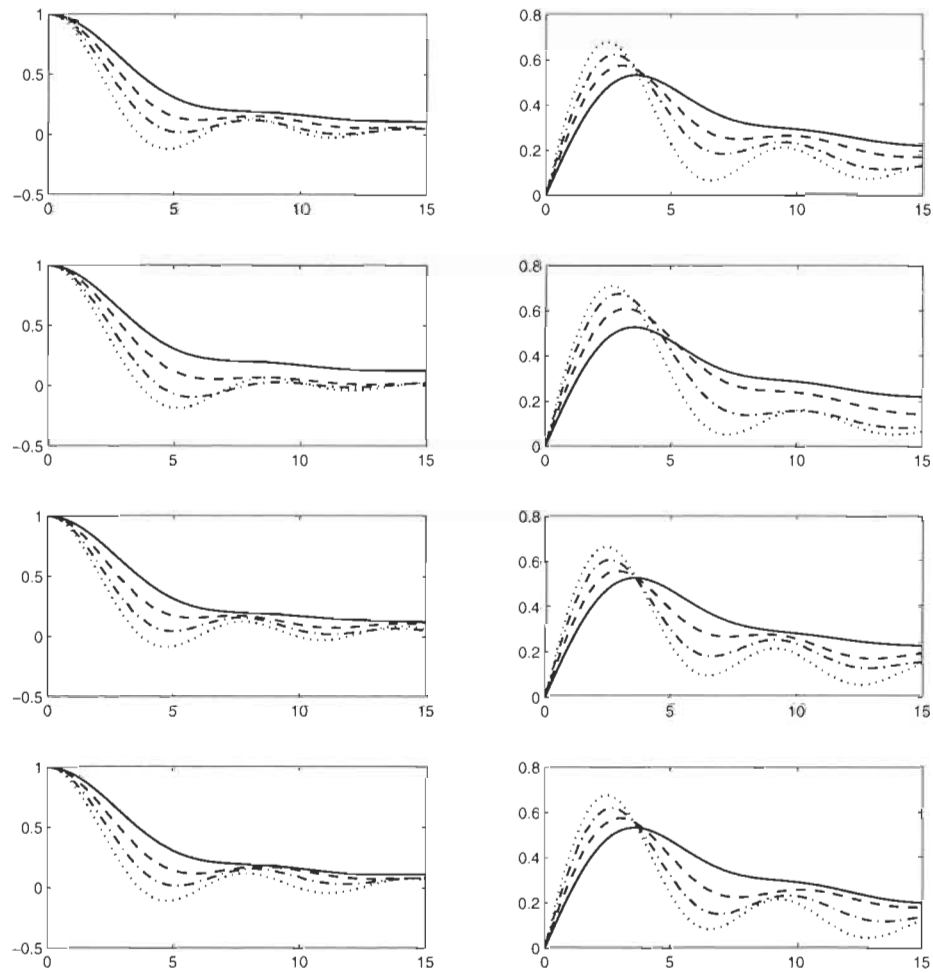


FIGURE 3.1 – Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ en fonction de $t \in [0, 15]$ pour les copules Normale (ligne 1), Clayton (ligne 2), Gumbel (ligne 3) et Plackett (ligne 4) pour $\tau(C) = 0$ (—), $\tau(C) = .25$ (---), $\tau(C) = .50$ (-·-) et $\tau(C) = .75$ (···).

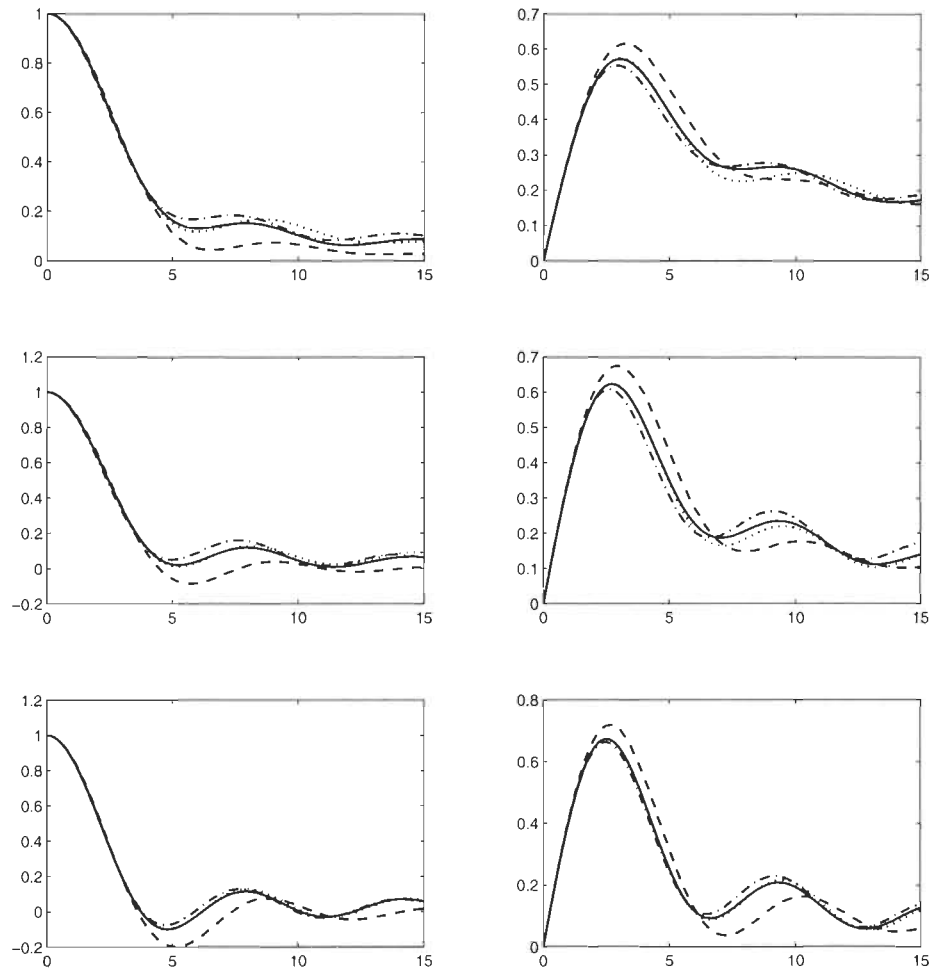


FIGURE 3.2 – Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ en fonction de $t \in [0, 15]$ pour $\tau(C) = .25$ (ligne 1), $\tau(C) = .50$ (ligne 2) et $\tau(C) = .75$ (ligne 3) pour les copules Normale (—), Clayton (---), Gumbel (-·-) et Plackett (···).

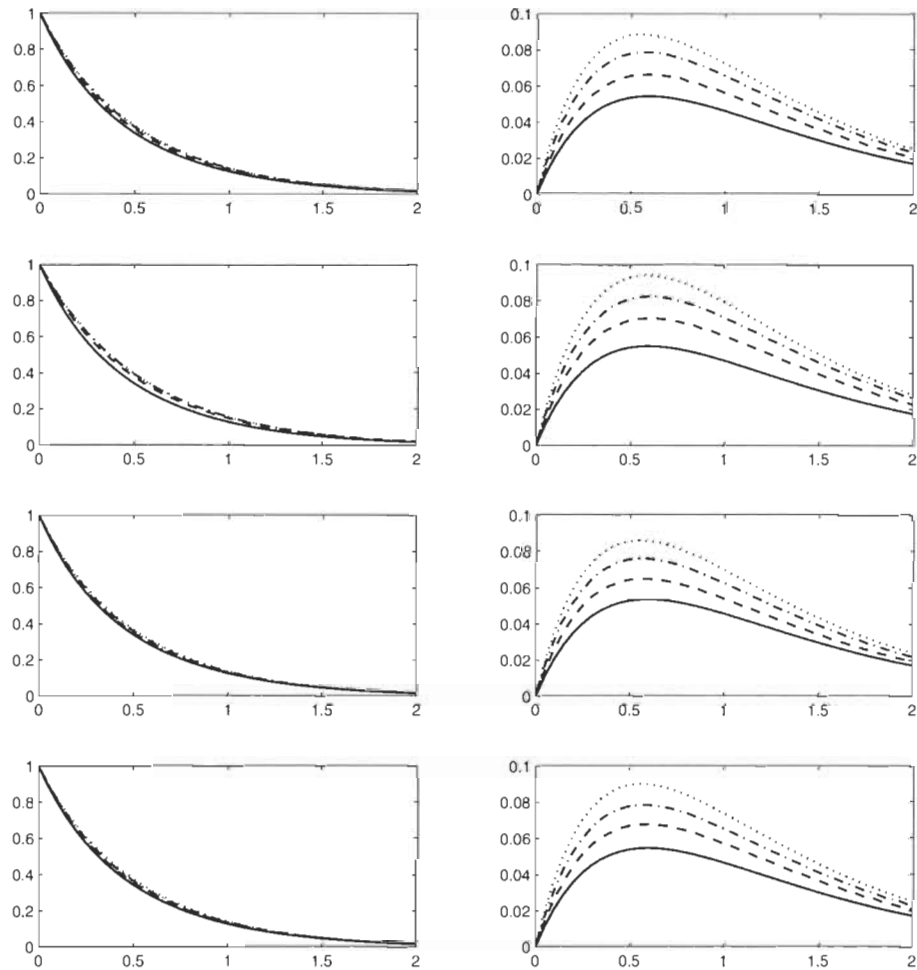


FIGURE 3.3 – Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W^\lambda(t)$ quand $\lambda = 1$ en fonction de $t \in [0, 2]$ pour les copules Normale (ligne 1), Clayton (ligne 2), Gumbel (ligne 3) et Plackett (ligne 4) pour $\tau(C) = 0$ (—), $\tau(C) = .25$ (---), $\tau(C) = .50$ (-·-) et $\tau(C) = .75$ (···).

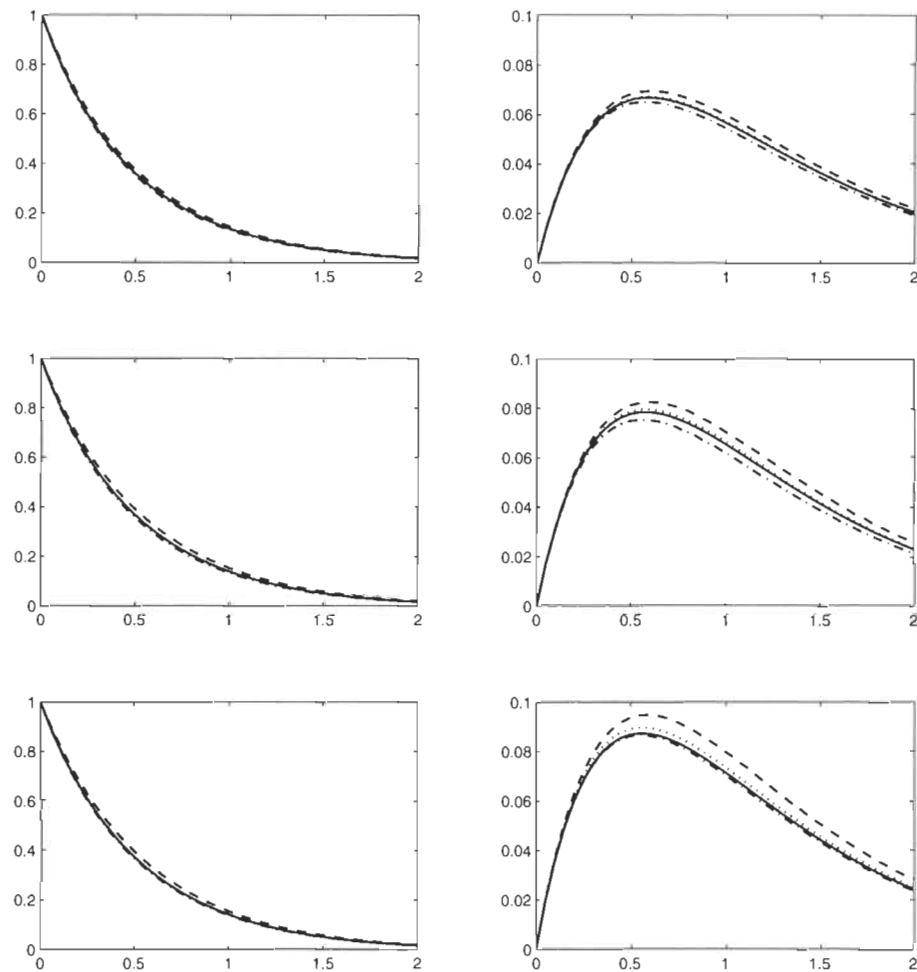


FIGURE 3.4 – Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W^\lambda(t)$ quand $\lambda = 1$ en fonction de $t \in [0, 2]$ pour $\tau(C) = .25$ (ligne 1), $\tau(C) = .50$ (ligne 2) et $\tau(C) = .75$ (ligne 3) pour les copules Normale (—), Clayton (---), Gumbel (-.) et Plackett (...).

À noter que la définition de la statistique S_n^g requiert une fonction de poids $g : \mathbb{R}^d \rightarrow \mathbb{R}$ qui est non-négative et intégrable ; cette fonction de poids est nécessaire pour que l'intégrale sur \mathbb{R} existe. En revanche, T_n^λ ne requiert pas de pondération supplémentaire, car ψ^λ est déjà pondérée dans sa définition ; cela libère donc du choix de g , qui peut être vu, d'une certaine façon, comme arbitraire.

Pour les mêmes raisons qu'il est suggéré de prendre la version estimée $V_{n,N}$ plutôt que V_n , on considérera des versions estimées de S_n^g et de T_n^λ , à savoir

$$S_{n,N}^g = n \int_{\mathbb{R}} |\psi_n(t) - \psi_N(t)|^2 g(t) dt \quad \text{et} \quad T_{n,N}^\lambda = n \int_{\mathbb{R}} |\psi_n^\lambda(t) - \psi_N^\lambda(t)|^2 dt,$$

où ψ_N et ψ_N^λ sont respectivement les estimations de ψ_θ et ψ_θ^λ basées sur les observations simulées $\mathbf{X}_1^*, \dots, \mathbf{X}_N^*$ i.i.d. $C_{\hat{\theta}_n}$. On peut déduire des formes explicites et facilement calculables pour $S_{n,N}^g$ et $T_{n,N}^\lambda$. En effet, il est montré à la Section 3.5 que

$$\begin{aligned} S_{n,N}^g &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \beta_g(\widehat{W}_j - \widehat{W}_k) \\ &\quad - \frac{2}{N} \sum_{j=1}^n \sum_{k=1}^N \beta_g(\widehat{W}_j - \widehat{W}_k^*) \\ &\quad + \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \beta_g(\widehat{W}_j^* - \widehat{W}_k^*), \end{aligned}$$

où la fonction $\beta_g(w) = \int_{\mathbb{R}} \cos(tw) g(t) dt$. Quand la fonction g est basée sur la densité Normale, c'est-à-dire que $g(t) \propto \phi(t/\lambda)$, on a

$$\beta_g(w) \propto \frac{\lambda}{2\sqrt{\pi}} \exp\left(-\frac{\lambda^2 w^2}{4}\right).$$

Dans ce cas, la statistique $S_{n,N}^g$ devient, à une constante multiplicative près,

$$\begin{aligned} S_{n,N}^\lambda &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \exp \left\{ -\frac{\lambda^2}{4} (\widehat{W}_j - \widehat{W}_k)^2 \right\} \\ &\quad - \frac{2}{N} \sum_{j=1}^n \sum_{k=1}^N \exp \left\{ -\frac{\lambda^2}{4} (\widehat{W}_j - \widehat{W}_k^*)^2 \right\} \\ &\quad + \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \exp \left\{ -\frac{\lambda^2}{4} (\widehat{W}_j^* - \widehat{W}_k^*)^2 \right\}. \end{aligned}$$

De même, en posant $\widehat{\alpha}_j = \widehat{W}_j(1 - \widehat{W}_j)$ et $\widehat{\alpha}_j^* = \widehat{W}_j^*(1 - \widehat{W}_j^*)$, on montre que

$$\begin{aligned} T_{n,N}^\lambda &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \beta^\lambda \left(\widehat{\alpha}_j \widehat{\alpha}_k, \widehat{W}_j - \widehat{W}_k \right) \\ &\quad - \frac{2}{N} \sum_{j=1}^n \sum_{k=1}^N \beta^\lambda \left(\widehat{\alpha}_j \widehat{\alpha}_k^*, \widehat{W}_j - \widehat{W}_k^* \right) \\ &\quad + \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \beta^\lambda \left(\widehat{\alpha}_j^* \widehat{\alpha}_k^*, \widehat{W}_j^* - \widehat{W}_k^* \right), \end{aligned}$$

où

$$\beta^\lambda(a, w) = \int_{-\infty}^{\infty} a^{\lambda|t|} \cos(tw) dt = \frac{-2\lambda \log(a)}{w^2 + \lambda^2 \log^2(a)}.$$

Pour davantage de détails concernant ce calcul, se référer à Meintanis *et al.* [52].

3.4.3 Comportement asymptotique de $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ sous \mathbb{H}_0

L'objectif de cette section est de décrire le comportement des statistiques de test $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ sous l'hypothèse nulle lorsque $n \rightarrow \infty$. Pour ce faire, on obtiendra d'abord le résultat pour S_n^g et T_n^λ . À cette fin, on définit les processus empiriques complexes $\Psi_n = \sqrt{n}(\psi_n - \psi_{\theta_n})$ et $\Psi_n^\lambda = \sqrt{n}(\psi_n^\lambda - \psi_{\theta_n}^\lambda)$, de telle sorte que l'on puisse écrire

$$S_n^g = \int_{\mathbb{R}} |\Psi_n(t)|^2 g(t) dt \quad \text{et} \quad T_n^\lambda = \int_{\mathbb{R}} |\Psi_n^\lambda(t)|^2 dt. \quad (3.7)$$

Il est utile de noter que l'on peut représenter ces statistiques en fonction du processus empirique $\tilde{\mathbb{K}}_n = \sqrt{n}(K_n - K_{\theta_n})$ utilisé pour l'adéquation à une copule via la statistique V_n^K . En effet, par une simple intégration par partie, on obtient

$$\begin{aligned}\psi_n(t) &= \int_0^1 e^{itw} dK_n(w) = e^{it} - it \int_0^1 e^{itw} K_n(w) dw, \\ \psi_{\theta_n}(t) &= \int_0^1 e^{itw} dK_{\theta_n}(w) = e^{it} - it \int_0^1 e^{itw} K_{\theta_n}(w) dw,\end{aligned}$$

de telle sorte que

$$\Psi_n(t) = -it \int_0^1 e^{itw} \tilde{\mathbb{K}}_n(w) dw.$$

Par des manipulations similaires, on montre que pour $q_t(w) = \{w(1-w)\}^{\lambda|t|} e^{itw}$

$$\Psi_n^\lambda(t) = - \int_0^1 \tilde{\mathbb{K}}_n(w) q'_t(w) dw.$$

Il n'est donc pas étonnant que les limites de $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ s'écrivent comme des fonctionnelles de la limite $\tilde{\mathbb{K}}$ de $\tilde{\mathbb{K}}_n$; ce résultat est énoncé formellement dans la proposition suivante; sa démonstration est reportée à la section 3.6.

Proposition 3.1. *Sous les hypothèses I-II, la statistique $S_{n,N}^g$ converge en loi sous \mathbb{H}_0 vers une limite \mathbb{S}^g de représentation*

$$\mathbb{S}^g = \int_0^1 \int_0^1 \tilde{\mathbb{K}}(w_1) \tilde{\mathbb{K}}(w_2) \tilde{\beta}_g(w_1 - w_2) dw_1 dw_2,$$

où $\tilde{\beta}_g(w) = \int_{\mathbb{R}} t^2 \cos(tw) g(t) dt$. De même, $T_{n,N}^\lambda$ converge en loi vers

$$\mathbb{T}^\lambda = \int_{\mathbb{R}} \left| \int_0^1 \tilde{\mathbb{K}}(w) q'_t(w) dw \right|^2 dt.$$

Les limites \mathbb{S}^g et \mathbb{T}^λ ne sont d'aucune utilité pour l'obtention des valeurs critiques des tests basés sur $S_{n,N}^g$ et $T_{n,N}^\lambda$. Ainsi, à l'instar du test basé sur $V_{n,N}^K$, on utilisera à cette fin le bootstrap paramétrique.

3.5 Dérivation de formules explicites pour les statistiques de test

3.5.1 Formule explicite pour $V_{n,N}^K$

En développant le carré à l'intérieur de l'intégrale dans la définition de $V_{n,N}^K$ à l'Équation (3.5), on obtient $V_{n,N}^K = A - 2B + C$, où

$$A = n \int_0^1 \{K_n(t)\}^2 dt, \quad B = n \int_0^1 K_n(t) K_N(t) dt \quad \text{et} \quad C = n \int_0^1 \{K_N(t)\}^2 dt.$$

En substituant la définition de K_n donnée à l'Équation (3.3), on obtient

$$\begin{aligned} A &= n \int_0^1 \left\{ \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \mathbf{1}(\widehat{W}_j \leq t) \mathbf{1}(\widehat{W}_k \leq t) \right\} dt \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \int_0^1 \mathbf{1}\{t \geq \max(\widehat{W}_j, \widehat{W}_k)\} dt \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \int_{\max(\widehat{W}_j, \widehat{W}_k)}^1 dt \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \left\{ 1 - \max(\widehat{W}_j, \widehat{W}_k) \right\}. \end{aligned}$$

Par des calculs semblables, on obtient

$$B = \frac{1}{N} \sum_{j=1}^n \sum_{k=1}^N \left\{ 1 - \max(\widehat{W}_j, \widehat{W}_k^*) \right\} \quad \text{et} \quad C = \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \left\{ 1 - \max(\widehat{W}_j^*, \widehat{W}_k^*) \right\}.$$

L'expression annoncée pour $V_{n,N}^K$ s'obtient finalement en rassemblant les trois termes.

3.5.2 Formule explicite pour $S_{n,N}^g$

En employant la formule d'Euler $e^{i\theta} = \cos(\theta) + i \sin(\theta)$, on obtient d'abord

$$\operatorname{Re}(\psi_n(t)) = \frac{1}{n} \sum_{j=1}^n \cos(t\widehat{W}_j) \text{ et } \operatorname{Im}(\psi_n(t)) = \frac{1}{n} \sum_{j=1}^n \sin(t\widehat{W}_j).$$

De là,

$$\begin{aligned} |\psi_n(t) - \psi_N(t)|^2 &= \left\{ \frac{1}{n} \sum_{j=1}^n \cos(t\widehat{W}_j) - \frac{1}{N} \sum_{j=1}^N \cos(t\widehat{W}_j^*) \right\}^2 \\ &\quad + \left\{ \frac{1}{n} \sum_{j=1}^n \sin(t\widehat{W}_j) - \frac{1}{N} \sum_{j=1}^N \sin(t\widehat{W}_j^*) \right\}^2 \\ &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \left\{ \cos(t\widehat{W}_j) \cos(t\widehat{W}_k) + \sin(t\widehat{W}_j) \sin(t\widehat{W}_k) \right\} \\ &\quad - \frac{2}{nN} \sum_{j=1}^n \sum_{k=1}^N \left\{ \cos(t\widehat{W}_j) \cos(t\widehat{W}_k^*) + \sin(t\widehat{W}_j) \sin(t\widehat{W}_k^*) \right\} \\ &\quad + \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \left\{ \cos(t\widehat{W}_j^*) \cos(t\widehat{W}_k^*) + \sin(t\widehat{W}_j^*) \sin(t\widehat{W}_k^*) \right\}. \end{aligned}$$

En utilisant par la suite l'identité trigonométrique $\cos(a)\cos(b) + \sin(a)\sin(b) = \cos(a-b)$, il vient

$$\begin{aligned} |\psi_n(t) - \psi_N(t)|^2 &= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n \cos \left\{ t \left(\widehat{W}_j - \widehat{W}_k \right) \right\} \\ &\quad - \frac{2}{nN} \sum_{j=1}^n \sum_{k=1}^N \cos \left\{ t \left(\widehat{W}_j - \widehat{W}_k^* \right) \right\} \\ &\quad + \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N \cos \left\{ t \left(\widehat{W}_j^* - \widehat{W}_k^* \right) \right\}. \end{aligned}$$

Pour compléter le calcul, il s'agit d'intégrer de chaque côté de la dernière expression par rapport à $dg(t)$ pour $t \in \mathbb{R}$.

3.5.3 Formule explicite pour $T_{n,N}^\lambda$

En procédant de façon similaire à celle employée pour $S_{n,N}^g$, on obtient

$$\begin{aligned}
|\psi_n^\lambda(t) - \psi_N^\lambda(t)|^2 &= \left\{ \frac{1}{n} \sum_{j=1}^n (\hat{\alpha}_j)^{\lambda|t|} \cos(t\widehat{W}_j) - \frac{1}{N} \sum_{j=1}^N (\hat{\alpha}_j^*)^{\lambda|t|} \cos(t\widehat{W}_j^*) \right\}^2 \\
&\quad + \left\{ \frac{1}{n} \sum_{j=1}^n (\hat{\alpha}_j)^{\lambda|t|} \sin(t\widehat{W}_j) - \frac{1}{N} \sum_{j=1}^N (\hat{\alpha}_j^*)^{\lambda|t|} \sin(t\widehat{W}_j^*) \right\}^2 \\
&= \frac{1}{n^2} \sum_{j=1}^n \sum_{k=1}^n (\hat{\alpha}_j \hat{\alpha}_k)^{\lambda|t|} \cos\left\{t(\widehat{W}_j - \widehat{W}_k)\right\} \\
&\quad - \frac{2}{nN} \sum_{j=1}^n \sum_{k=1}^N (\hat{\alpha}_j \hat{\alpha}_k^*)^{\lambda|t|} \cos\left\{t(\widehat{W}_j - \widehat{W}_k^*)\right\} \\
&\quad + \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N (\hat{\alpha}_j^* \hat{\alpha}_k^*)^{\lambda|t|} \cos\left\{t(\widehat{W}_j^* - \widehat{W}_k^*)\right\}.
\end{aligned}$$

On obtient la forme explicite annoncée pour $T_{n,N}^\lambda$ en intégrant de chaque côté de cette expression par rapport à dt pour $t \in \mathbb{R}$.

3.6 Preuve de la Proposition 3.1

Par la forme de S_n^g obtenue à l'Équation (3.7), on a

$$S_n^g = \int_{\mathbb{R}} \left| -it \int_0^1 e^{itw} \widetilde{\mathbb{K}}_n(w) dw \right|^2 g(t) dt.$$

Par une application directe du Théorème des applications continues, on déduit que S_n^g converge nécessairement en loi vers \mathbb{S}^g . En développant, on obtient

$$\begin{aligned}\mathbb{S}^g &= \int_{\mathbb{R}} \left| -it \int_0^1 e^{itw} \tilde{\mathbb{K}}(w) dw \right|^2 g(t) dt \\ &= \int_{\mathbb{R}} t^2 \left| \int_0^1 \sin(tw) \tilde{\mathbb{K}}(w) dw - i \int_0^1 \cos(tw) \tilde{\mathbb{K}}(w) dw \right|^2 g(t) dt.\end{aligned}$$

Comme $|z|^2 = |a + bi|^2 = a^2 + b^2$, on trouve

$$\begin{aligned}\mathbb{S}^g &= \int_{\mathbb{R}} \left\{ \left(\int_0^1 \sin(tw) \tilde{\mathbb{K}}(w) dw \right)^2 + \left(\int_0^1 \cos(tw) \tilde{\mathbb{K}}(w) dw \right)^2 \right\} t^2 g(t) dt \\ &= \int_{\mathbb{R}} \left[\int_0^1 \int_0^1 \{ \sin(tw_1) \sin(tw_2) + \cos(tw_1) \cos(tw_2) \} \tilde{\mathbb{K}}(w_1) \tilde{\mathbb{K}}(w_2) dw_1 dw_2 \right] t^2 g(t) dt.\end{aligned}$$

En utilisant l'identité trigonométrique $\cos(a) \cos(b) + \sin(a) \sin(b) = \cos(a - b)$,

$$\begin{aligned}\mathbb{S}^g &= \int_{\mathbb{R}} \left[\int_0^1 \int_0^1 \cos \{t(w_1 - w_2)\} \tilde{\mathbb{K}}(w_1) \tilde{\mathbb{K}}(w_2) dw_1 dw_2 \right] t^2 g(t) dt \\ &= \int_0^1 \int_0^1 \tilde{\mathbb{K}}(w_1) \tilde{\mathbb{K}}(w_2) \tilde{\beta}_g(w_1 - w_2) dw_1 dw_2.\end{aligned}$$

Enfin, on note que

$$T_n^\lambda = \int_{\mathbb{R}} \left| \int_0^1 \tilde{\mathbb{K}}_n(w) q'_t(w) dw \right|^2 dt.$$

Par le Théorème des applications continues, on déduit que T_n^λ converge en loi vers

$$\mathbb{T}^\lambda = \int_{\mathbb{R}} \left| \int_0^1 \tilde{\mathbb{K}}(w) q'_t(w) dw \right|^2 dt.$$

Chapitre 4

Les nouveaux tests en action

4.1 Préliminaires sur les simulations

4.1.1 Justification de l'usage d'une méthode Monte–Carlo

Maintenant que nous avons développé les bases théoriques de nos nouveaux tests et obtenu des formes explicites pour $S_{n,N}^g$ et $T_{n,N}^\lambda$, nous sommes en mesure de tester la validité de ceux-ci à l'aide de simulations. De manière générale, il y a deux caractéristiques à considérer lors de l'application d'un test statistique pour confronter des hypothèses \mathbb{H}_0 et \mathbb{H}_1 :

- (i) Son erreur de type I, c'est-à-dire $\alpha = \mathbb{P}(\text{rejeter } \mathbb{H}_0 | \mathbb{H}_0 \text{ est vraie})$.
- (ii) Sa puissance, c'est-à-dire $\mathcal{P} = \mathbb{P}(\text{rejeter } \mathbb{H}_0 | \mathbb{H}_0 \text{ est fausse})$.

Dans le cas de tests simples, comme les tests sur une ou deux moyennes basés sur la distribution normale ou Student, il est facile de définir un critère de rejet qui fait en sorte que l'erreur de type I corresponde *exactement* au seuil α désiré ; il est aussi possible dans ces situations d'obtenir \mathcal{P} sous forme de courbe de puissance sous une

famille adéquatement définie de contre-hypothèses à \mathbb{H}_0 . Dans le cas des tests basés sur $S_{n,N}^\lambda$, $T_{n,N}^\lambda$ et $V_{n,N}^K$, de tels calculs ne sont malheureusement pas possibles. La façon de s'en sortir dans de tels cas est de procéder par simulations : on génère un grand nombre d'échantillons selon une certaine loi et la probabilité de rejet du test est estimée par la proportion empirique. Il s'agit d'un cas particulier des méthodes de simulation dites *Monte-Carlo*.

4.1.2 Statistiques de test étudiées

Les statistiques de test qui seront étudiées dans les simulations sont $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ pour $N = 250$ et $\lambda \in \{1/2, 1, 3, 5\}$. Dans la suite, on notera ces statistiques

$$S_{n,N}^{\lambda=1/2}, \quad S_{n,N}^{\lambda=1}, \quad S_{n,N}^{\lambda=3}, \quad S_{n,N}^{\lambda=5}, \quad T_{n,N}^{\lambda=1/2}, \quad T_{n,N}^{\lambda=1}, \quad T_{n,N}^{\lambda=3}, \quad T_{n,N}^{\lambda=5}.$$

On comparera l'efficacité des tests basés sur ces statistiques à deux tests d'adéquation concurrents et très populaires. Le premier est le test de Genest *et al.* [33] basé sur la statistique $V_{n,N}^K$. Le deuxième est un test décrit par Genest *et al.* [37] basé sur une statistique notée $V_{n,N}^C$ qui est calculée à partir de la copule empirique ; les détails sur ce test se retrouvent à l'Annexe A.3. Par soucis de concision, les résultats présentés se concentrent uniquement sur le cas bidimensionnel, c'est-à-dire quand $d = 2$. Également, on testera des modèles dont le paramètre inconnu est unidimensionnel ; dans ce cas, $\theta \in \mathbb{R}$ est facilement estimé par la méthode d'inversion du tau de Kendall, telle que décrite à l'Annexe A.2.

4.1.3 Stratégie pour l'obtention des valeurs critiques

Pour faciliter la présentation des résultats, les modèles de copules considérés sous l'hypothèse nulle \mathbb{H}_0 seront restreints aux familles Normale, Clayton et Gumbel ; ces

modèles de dépendance sont présentés au Chapitre 2. Dans une étude de simulations, le nombre de répétitions nécessaires pour obtenir des estimations fiables des probabilités de rejet doit être élevé (par exemple, 1 000 répétitions). Dans le cas des tests basés sur les statistiques $S_{n,N}^\lambda$, $T_{n,N}^\lambda$, $V_{n,N}^K$ et $V_{n,N}^C$, chacune de ces répétitions doit à son tour mettre en marche un grand nombre de répétitions de la méthode du bootstrap paramétrique. Ainsi, le temps d'exécution d'une telle démarche est astronomique ! Il est donc impossible d'envisager une telle possibilité pour cette étude.

Pour contourner ce problème et ainsi être en mesure d'étudier les statistiques de test sous plusieurs scénarios intéressants, on a construit des tableaux de valeurs critiques pour les trois familles de copules considérées sous \mathbb{H}_0 , c'est-à-dire Normale, Clayton et Gumbel. Le Tableau B.1 présenté en annexe montre les valeurs critiques de rejet d'une copule Normale au seuil $\alpha = 0,05$ (*i.e.* le 95ème percentile) en fonction du tau de Kendall $\tau(C)$, estimées à l'aide de 1 000 répétitions pour les dix tests étudiés et lorsque la taille d'échantillon est $n \in \{100, 250\}$.

Une façon d'utiliser ces tableaux serait d'interpoler linéairement entre deux valeurs possibles. Ainsi, si la valeur du tau de Kendall pour un jeu de données est $\tau_n = 0,48$, la valeur critique du test serait obtenue en interpolation entre les valeurs critiques estimées lorsque $\tau(C) = 0,45$ et $\tau(C) = 0,50$. Cette façon de faire est assimilable à une version approximative de l'inversion du tau de Kendall ; une telle manière de chercher dans une liste peut cependant s'avérer fastidieuse, surtout lorsqu'elle est répétée un grand nombre de fois. Pour augmenter la vitesse d'exécution, on ajustera les valeurs critiques en fonction de $\tau(C)$ à l'aide de régressions polynomiales. En effet, l'usage de cette méthode permet certains avantages plutôt pratiques dont la simplicité des modèles polynomiaux obtenus. On obtiendra alors un modèle prédictif de la valeur critique VC en fonction de $\tau = \tau(C)$ de la forme

$$VC(\tau) = \beta_0 + \beta_1 \tau + \beta_2 \tau^2 + \dots + \beta_p \tau^p.$$

Pour chaque statistique, chaque taille d'échantillons $n \in \{100, 250\}$ et chacun des trois modèles de copules sous \mathbb{H}_0 , on a choisi le modèle adéquat entre $p = 2$ et $p = 3$ selon la valeur du R^2 . Quelques exemples de telles courbes ainsi obtenues et passant à travers les valeurs critiques se retrouvent à la Figure 4.1. Les courbes présentées sont respectivement (de gauche à droite et de bas en haut),

$$VC(\tau) = 0,0232 + 0,0001\tau - 0,0678\tau^2 + 0,0444\tau^3,$$

$$VC(\tau) = 0,0499 - 0,0036\tau - 0,1207\tau^2 + 0,0719\tau^3,$$

$$VC(\tau) = 0,0744 + 0,0327\tau - 0,2926\tau^2 + 0,1916\tau^3,$$

$$VC(\tau) = 0,0849 - 0,0347\tau - 0,0590\tau^2.$$

Par exemple, si on obtient $\tau_n = 0,48$ pour un jeu de données de taille $n = 100$, alors la valeur critique du test basé sur $S_{n,N}^{\lambda=1}$ pour tester la copule Normale est estimée à

$$VC = 0,0232 + 0,0001(0,48) - 0,0678(0,48)^2 + 0,0444(0,48)^3 = 0,0125.$$

On rejettera donc \mathbb{H}_0 si $S_{n,N}^{\lambda=1} > 0,0125$. La Figure B.1 en annexe superpose ces courbes pour les trois modèles de copules considérés sous \mathbb{H}_0 lorsque $n = 100$; la Figure B.2 est similaire, mais pour $n = 250$. En résumé, l'estimation d'une probabilité de rejet \mathcal{P} d'un test d'adéquation basé sur une statistique S_n pour tester une hypothèse nulle \mathbb{H}_0 lorsque le modèle est sous \mathbb{H}_1 se fera ainsi :

Étape 1. On simule un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ i.i.d. C ;

Étape 2. À partir de $(X_1, Y_1), \dots, (X_n, Y_n)$, on calcule la statistique de test S_n sous \mathbb{H}_0 , ainsi que le tau de Kendall empirique τ_n ;

Étape 3. On détermine la valeur critique du test à partir de la courbe des valeurs critiques, c'est-à-dire $VC(\tau_n)$;

Étape 4. On rejette \mathbb{H}_0 si $S_n > VC(\tau_n)$.

On répète les étapes 1–4 un nombre élevé B de fois, et \mathcal{P} est estimée par

$$\hat{\mathcal{P}} = \frac{1}{B} \sum_{\ell=1}^B \mathbf{1}(S_n > VC(\tau_n)).$$

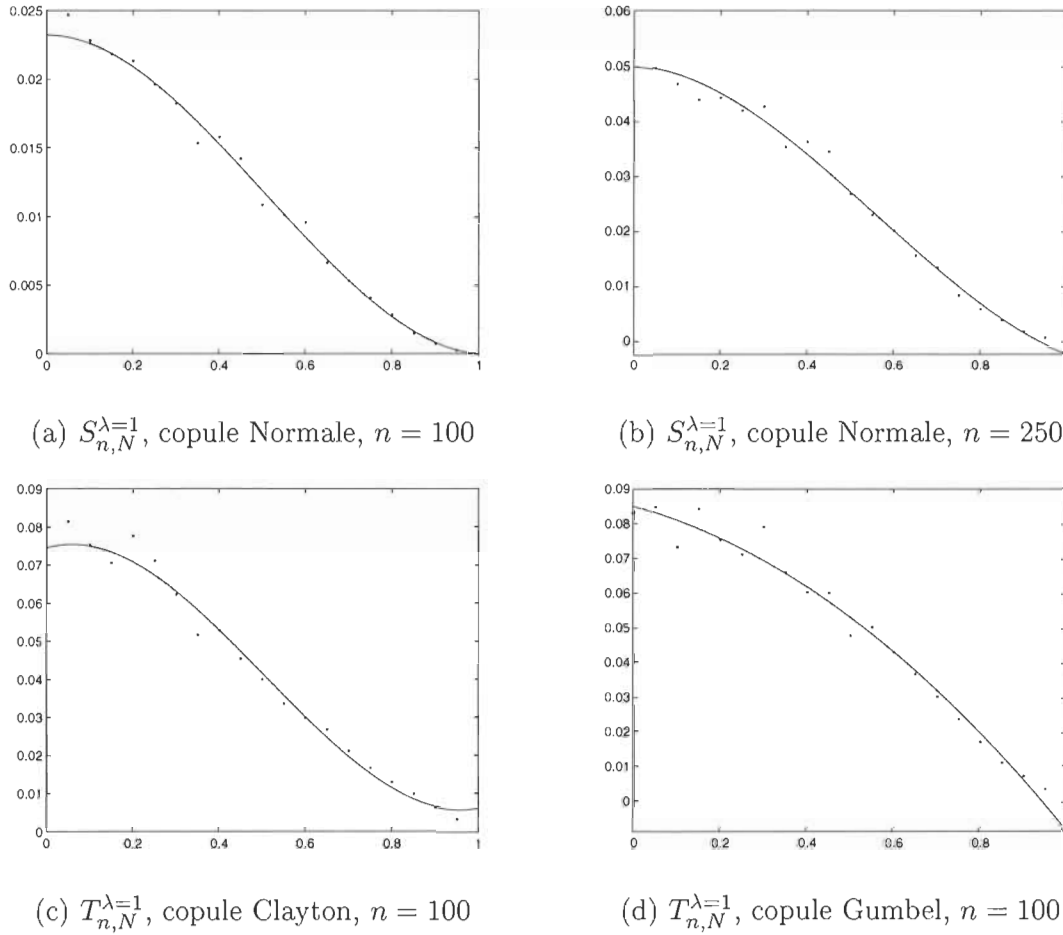


FIGURE 4.1 – Régressions polynomiales en fonction de τ des valeurs critiques pour les statistiques de test $S_{n,N}^{\lambda=1}$ et $T_{n,N}^{\lambda=1}$.

4.2 Capacité des tests à conserver leur seuil nominal

Avant d'étudier la puissance des tests, il faut s'assurer que ceux-ci conservent correctement leur seuil nominal $\alpha = 0,05$ sous \mathbb{H}_0 . En fait, l'incapacité d'un test à conserver ce seuil rendrait invalide d'éventuels résultats sur sa puissance. Le Tableau 4.3 montre les résultats de simulation obtenus lorsque la copule simulée sous \mathbb{H}_1 est la même que celle qui est testée ; on a considéré les copules Normale (N), Clayton (C ℓ) et Gumbel (Gu). Trois niveaux de dépendance tels que mesurés par le tau de Kendall ont été considérés, à savoir $\tau(C) \in \{.25, .50, .75\}$.

Selon les résultats du Tableau 4.3, on peut affirmer que les dix statistiques étudiées sont généralement excellentes pour maintenir leur seuil nominal. En effet, la majorité des valeurs se maintiennent dans l'intervalle $5\% \pm 1\%$. En analysant ces résultats, il faut garder à l'esprit que deux sources de variabilité les ont influencés. D'abord, les valeurs critiques ont été approximées par le bootstrap paramétrique. De surcroît, ces valeurs critiques ont été elles-mêmes ajustées à l'aide d'une régression polynomiale.

4.3 Étude sur la puissance des tests

La capacité de rejeter le plus souvent possible une hypothèse nulle lorsque celle-ci est fausse est une qualité primordiale pour un test statistique. Cette aptitude se nomme la puissance d'un test. Dans cette section, les probabilités de rejet de l'hypothèse nulle \mathbb{H}_0 sous plusieurs types de contre-hypothèses seront considérées pour les statistiques à l'étude. Autrement dit, on étudie la puissance des tests d'adéquation à une copule. Pour ce faire, outre les copules Normale, Clayton et Gumbel étudiées à la Section 4.2, on considère les copules de Plackett (Pl) et de Frank (Fr) sous \mathbb{H}_1 . À l'instar de la Section 4.2, les tailles d'échantillons considérées sont $n = \{100, 250\}$, et on a posé $N = 250$. Le Tableau 4.4 montre les résultats de puissance obtenus lorsque la copule sous \mathbb{H}_0 est la Normale ; les tableaux 4.5 et 4.6 concernent respectivement les hypothèses \mathbb{H}_0 d'une copule de Clayton et Gumbel.

Telles qu'attendues, les puissances sont supérieures lorsque $n = 250$, en comparaison de $n = 100$. On remarque aussi que de manière générale, les probabilités de rejet de l'hypothèse nulle sont plus élevées lorsque la valeur de $\tau(C)$ est élevée. Ceci s'explique par le fait que les modèles sont plus distinctifs pour des niveaux de dépendance élevés ; les tests ont alors plus de facilité à les distinguer. Il est intéressant de noter que les nouveaux tests basés sur $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ sont systématiquement plus puissants que ceux basés sur $V_{n,N}^K$ et $V_{n,N}^C$. En outre, la statistique de test $T_{n,N}^\lambda$ donne généralement de

meilleurs résultats que sa rivale $S_{n,N}^\lambda$. À noter toutefois que la performance de $S_{n,N}^\lambda$ est faible pour $\lambda = \{0.5, 1\}$. Ainsi, on peut affirmer que les tests basés sur $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$ sont remarquablement performants par rapport aux deux procédures populaires.

4.4 Compléments sur les tests de l'hypothèse d'une copule normale

Il serait intéressant d'étudier la puissance des tests d'adéquation pour la famille de copules normales sous des contre-hypothèses naturelles à ce modèle de dépendance. Parmi les possibilités, on a les copules elliptiques introduites par Fang *et al.* [22], qui sont extraites de la famille générale des lois elliptiques étudiées par Cambanis *et al.* [9]. On retrouve comme cas particuliers de cette famille générale les copules de Student (voir Demarta et McNeil [19]) et de Laplace généralisée (voir Kozubowski *et al.* [50]). Cependant, quelques expériences exploratoires montrent que la puissance des tests de normalité basés sur les nouvelles statistiques $S_{n,N}^\lambda$ et $T_{n,N}^\lambda$, de même que celle de $V_{n,N}^K$ et $V_{n,N}^C$, est très faible, c'est-à-dire des résultats à peine plus élevés que la valeur $\alpha = 0,05$ du seuil nominal.

Une autre possibilité consiste à utiliser les copules Khi-deux introduites par Bârdossy [8] et étudiées par Quessy *et al.* [61]. Cette famille de modèles de dépendance est décrite à la Section 2.4.6. Dans le cas bidimensionnel, la copule Khi-deux de corrélation ρ et de paramètre de décentralité $a \in [0, \infty)$ est la structure de dépendance extraite de la paire $((Z_1 + a)^2, (Z_2 + a)^2)$, où (Z_1, Z_2) est distribuée selon la loi normale bivariée de moyennes nulles, de variances unitaires et de coefficient de corrélation ρ . Comme cas particulier, on retrouve la copule Khi-deux centrée quand le paramètre a de décentralité est nul ; la copule normale apparaît comme cas limite lorsque $a \rightarrow \infty$.

La Figure 4.2 montre les courbes de puissance des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$ et $V_{n,N}^K$ en fonction du paramètre de décentralité $a \in [0, 1]$. Le choix de $S_{n,N}^{\lambda=3}$ et $T_{n,N}^{\lambda=3}$ repose sur le fait que ces statistiques se sont avérées les plus performantes pour tester la normalité. On a considéré les tailles d'échantillons $n = 100$ (figures de gauche) et $n = 250$ (figures de droite), de même que trois niveaux de dépendance associés à $\tau(C) \in \{.25, .50, .75\}$ qui correspondent respectivement aux figures du haut, du milieu et du bas.

Un premier regard à la Figure 4.2 montre ce qui était attendu : la puissance est supérieure lorsque $n = 250$. À cet effet, il est possible d'affirmer que la puissance sera encore meilleure si n est supérieur à $n = 250$. De plus, les probabilités de rejet de l'hypothèse nulle augmentent en fonction de la valeur de $\tau(C)$; ce phénomène a également été observé dans l'étude de puissance dont les résultats sont répertoriés aux tableaux 4.4–4.6.

Tel qu'attendu, le paramètre de décentralité a exerce une influence notable sur la puissance des tests. En fait, on voit que leur puissance est inversement proportionnelle à la valeur de a . Ce comportement s'explique facilement par la construction même de la copule Khi-deux : le cas $a \rightarrow \infty$ correspond à la copule Normale, alors que $a = 0$ fournit d'une certaine manière le modèle *le plus éloigné* de la copule Normale. À notre grand plaisir, on note que les nouveaux tests basés sur $S_{n,N}^{\lambda=3}$ et $T_{n,N}^{\lambda=3}$ sont systématiquement meilleurs que celui basé sur $V_{n,N}^K$ pour détecter des écarts de type Khi-deux, et ce pour l'ensemble des cas considérés. Enfin, comme le test qui utilise $S_{n,N}^{\lambda=3}$ est significativement plus puissant que son concurrent basé sur $T_{n,N}^{\lambda=3}$, on pourrait recommander l'usage de $S_{n,N}^{\lambda=3}$ pour tester une structure de dépendance normale.

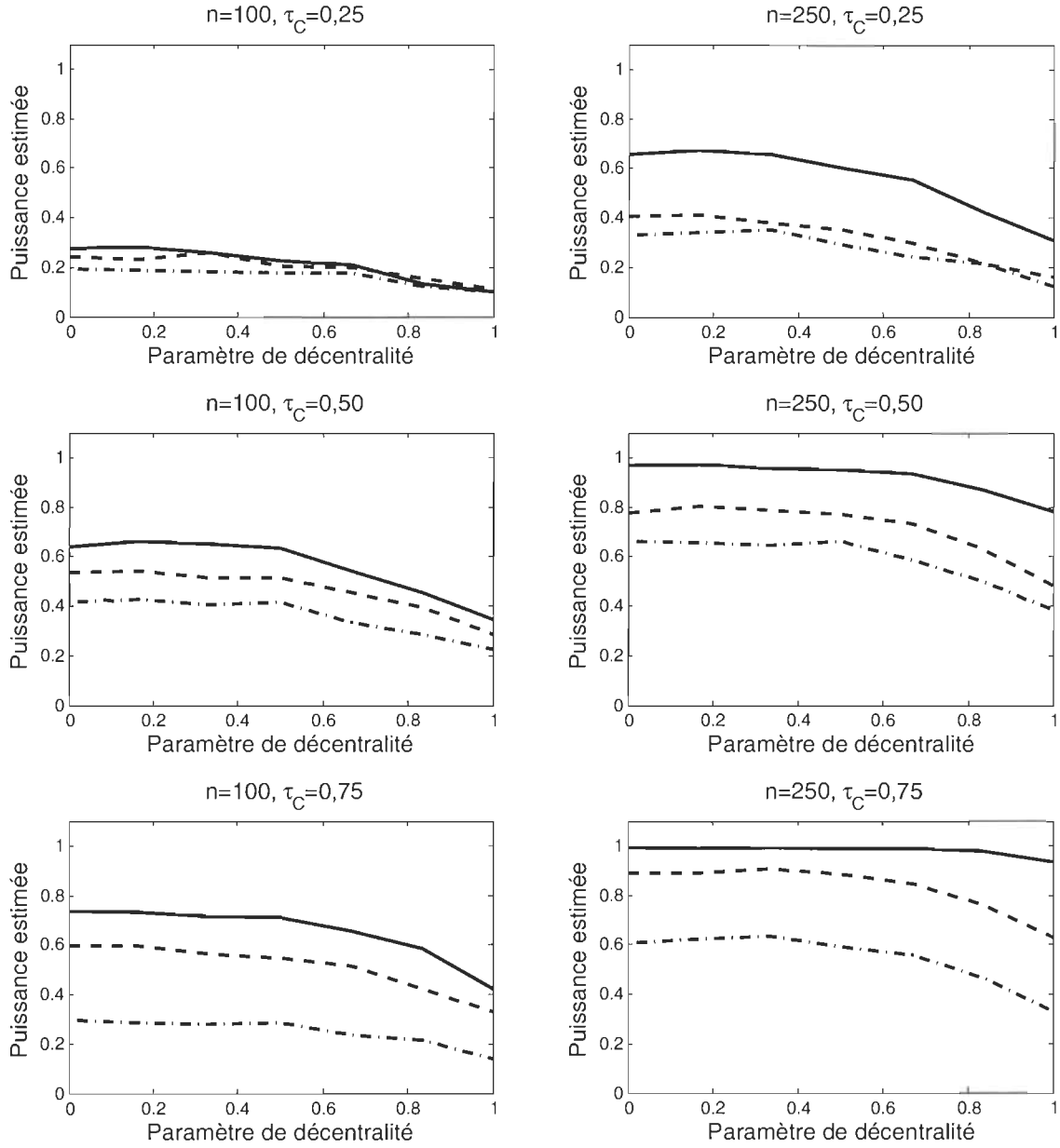


FIGURE 4.2 – Puissance des tests d'adéquation à la copule normale basés sur $S_{n,N}^{\lambda=3}$ (—), $T_{n,N}^{\lambda=3}$ (--) et $V_{n,N}^K$ (-·-) en fonction du paramètre de décentralité $a \in [0, 1]$ de la copule Khi-deux bidimensionnelle.

4.5 Illustration sur le jeu de données de Cook & Johnson (1986)

4.5.1 Présentation des données

Un jeu de données étudié à l'origine par Cook et Johnson [12] sera analysé ici dans le but de mettre en application nos tests statistiques proposés. Ces données consistent en $n = 655$ analyses chimiques provenant d'échantillons d'eau collectés dans la région de Montrose au Colorado et concernant les concentrations d'Uranium (U), de Lithium (Li), de Cobalt (Co), de Potassium (K), de Caesium (Cs), de Scandium (Sc) et de Titanium (Ti). Pour l'analyse qui sera présentée dans la suite, on s'est concentré sur les variables Co, Cs, Sc et Ti, car celles-ci fournissent les formes de dépendance les plus variées et intéressantes. On peut d'ailleurs se faire une idée des types de dépendance entre les paires en examinant la Figure 4.3. On peut aussi saisir la force du lien entre ces variables en calculant le tau de Kendall empirique, c'est-à-dire

$$\tau_n = \frac{4}{n(n-1)} \sum_{j < k} \mathbf{1} \{ (X_j - X_k)(Y_j - Y_k) > 0 \} - 1.$$

Pour les paires (Co, Cs), (Co, Sc), (Co, Ti), (Sc, Ti), (Cs, Sc) et (Cs, Ti), on a respectivement $\tau_n = 0,151$, $\tau_n = 0,515$, $\tau_n = 0,354$, $\tau_n = 0,218$, $\tau_n = 0,297$ et $\tau_n = 0,427$. En regardant les paires de rangs standardisés (les graphiques sous la diagonale à la Figure 4.3), il semble que la dépendance entre les variables soit symétrique par rapport à la seconde diagonale. Autrement dit, la copule entre chacune des paires semble posséder la propriété de symétrie radiale. Le but ultime ici est d'identifier formellement un modèle de dépendance adéquat pour le vecteur de variables (Co, Cs, Sc, Ti) à l'aide des procédures de test introduites dans ce travail. Pour ce faire, on débutera par une analyse paire-par-paire. Ensuite, on tentera de trouver un modèle à quatre dimensions approprié pour (Co, Cs, Sc, Ti) en appliquant les tests pour $d = 4$.

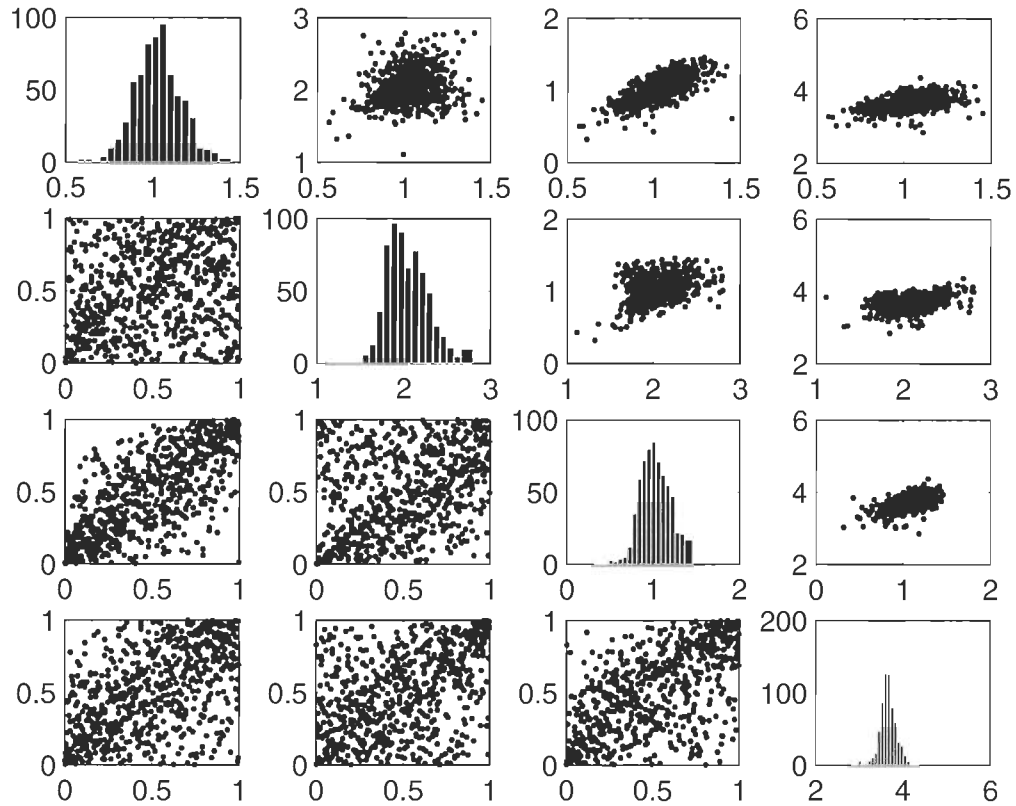


FIGURE 4.3 – Nuages de points (données brutes et rangs standardisés) et histogrammes pour les variables Co, Cs, Sc et Ti des données de Cook et Johnson [12].

4.5.2 Analyse paire par paire

Les tests d'adéquation basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$ et $V_{n,N}^K$ ont été effectués pour les six paires possibles de variables, c'est-à-dire (Co, Cs), (Co, Sc), (Co, Ti), (Sc, Ti), (Cs, Sc) et (Cs, Ti). On a testé quatorze structures de dépendance sous l'hypothèse nulle, à savoir les copules Normale (N), Khi-deux centrée (χ^2), Student à $\nu \in \{1, 5\}$ degrés de liberté (T_1 , T_5), Fisher à $\nu \in \{1, 5\}$ degrés de liberté (F_1 , F_5), Plackett (Pl), Clayton (Cl), Frank (Fr) et Gumbel (Gu). On a aussi considéré les versions *Squared* des copules de Plackett, Clayton, Frank et Gumbel (notées respectivement P_ℓ^2 , Cl^2 , Fr^2 et Gu^2) décrites par Quessy et Durocher [60](2019). Pour faire court, la version *Squared* d'une copule C est la structure de dépendance de $(|2U_1 - 1|, |2U_2 - 1|)$ pour $(U_1, U_2) \sim C$. Les copules Khi-deux et Fisher sont des cas particuliers de cette construction.

Les résultats complets de ces analyses sont présentés aux Tableaux B.2, B.3 et B.4; le code a été exécuté avec $N = 250$ et $M = 500$ répétitions pour la méthode de bootstrap paramétrique. Il a fallu une cinquantaine d'heures pour estimer toutes ces p-valeurs. Le Tableau 4.1 offre une synthèse des résultats obtenus en identifiant les modèles de copules qui ont été acceptés et ceux qui ont été rejetés. Ici, on a considéré qu'un modèle est rejeté dès lors que le test basé sur $S_{n,N}^{\lambda=3}$ ou le test basé sur $T_{n,N}^{\lambda=3}$ rejette \mathbb{H}_0 .

À la lumière des résultats du Tableau 4.1, on dénote une tendance assez nette à accepter les modèles qui possèdent la propriété de symétrie radiale. En effet, sauf dans le cas de la paire (Cs, Ti), les copules Normale, Student à $\nu = 5$ degrés de liberté et Plackett sont acceptés; on remarque aussi que la copule de Student à $\nu = 1$ degré de liberté est accepté pour quatre des six paires. Cette constatation permet d'entrevoir qu'un modèle adéquat à $d = 4$ dimensions pour le vecteur (Co, Cs, Sc, Ti) devra posséder la symétrie radiale; ce choix formel est l'objet de la section suivante.

TABLE 4.1 – Tableau-synthèse des modèles de copules acceptés et rejetés pour six paires de variables des données de Cook et Johnson [12].

Paire	Modèles de copules acceptés	Modèles de copules refusés
(Co,Cs)	N, T_1 , T_5 , Pl, Fr	χ^2 , F_1 , F_5 , Pl^2 , Cl , Cl^2 , Fr^2 , Gu, Gu^2
(Co,Sc)	N, T_1 , T_5 , Pl, Fr, Fr^2	χ^2 , F_1 , F_5 , Pl^2 , Cl , Cl^2 , Gu, Gu^2
(Co,Ti)	N, T_1 , T_5 , Pl, Fr, Gu	χ^2 , F_1 , F_5 , Pl^2 , Cl , Cl^2 , Fr^2 , Gu^2
(Sc,Ti)	N, T_1 , T_5 , Pl, Pl^2 , Cl^2	χ^2 , F_1 , F_5 , Cl , Fr, Fr^2 , Gu, Gu^2
(Cs,Sc)	N, T_5 , Pl, Cl , Fr^2	χ^2 , T_1 , F_1 , F_5 , Pl^2 , Cl^2 , Fr, Gu, Gu^2
(Cs,Ti)	χ^2 , F_1 , F_5 , Pl^2 , Cl^2 , Gu, Gu^2	N, T_1 , T_5 , Pl, Cl , Fr^2 , Fr

Pour compléter, les graphiques supérieures de la Figure 4.4 présentent les courbes de ψ_n^{Re} et de ψ_n^{Im} pour la paire (Co, Sc); on retrouve $\psi_n^{\lambda=3,\text{Re}}$ et de $\psi_n^{\lambda=3,\text{Im}}$ au bas

de cette même Figure 4.4. Sur ces graphiques, on a superposé les versions théoriques sous les copules Normale (modèle accepté) et Gumbel (modèle rejeté). On constate que les versions empiriques sont assez près des versions théoriques sous l'hypothèse d'une copule Normale, ce qui explique l'acceptation de ce modèle. En revanche, les écarts des versions empiriques avec les courbes sous la copule de Gumbel sont assez importants, ce qui est cohérent avec le rejet de ce modèle de dépendance.

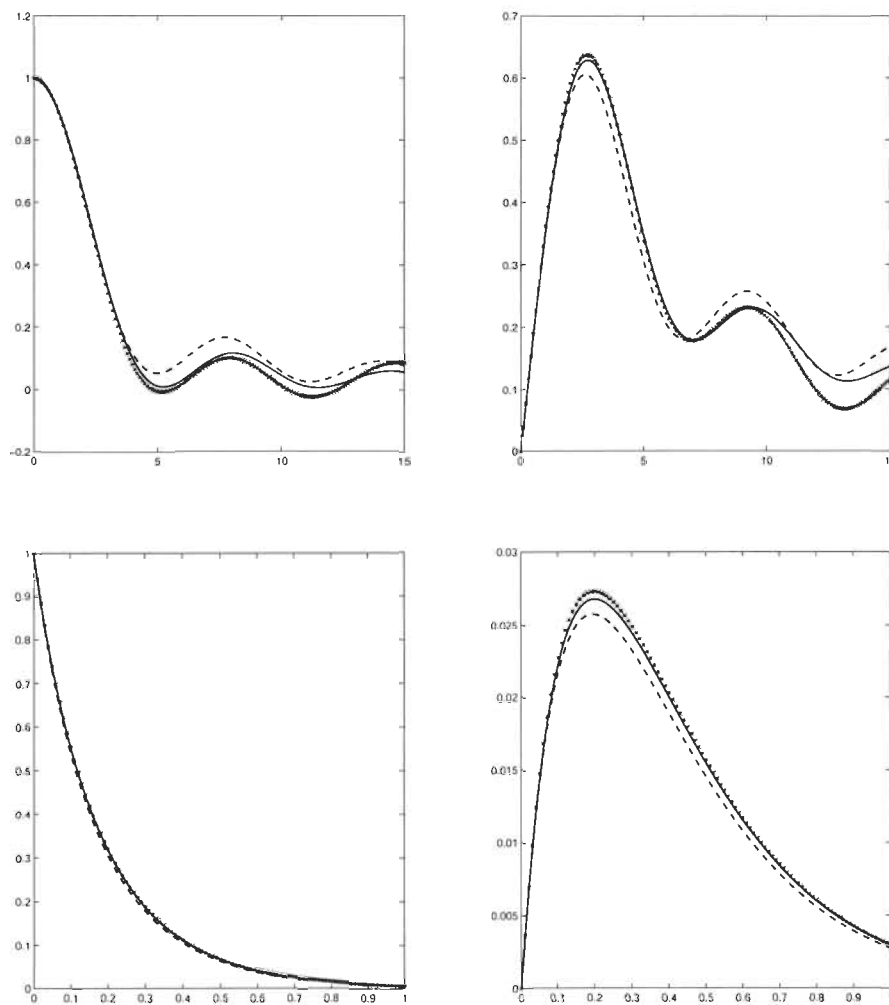


FIGURE 4.4 – Parties réelles (figures de gauche) et imaginaires (figures de droite) de $\psi_W(t)$ et $\psi_n(t)$ (***) en fonction de $t \in [0, 15]$ (haut) et de $\psi_W^\lambda(t)$ et $\psi_n^\lambda(t)$ (***) quand $\lambda = 3$ en fonction de $t \in [0, 1]$ (bas) pour un $\tau(C)$ fixé pour les copules Normale (—) et Gumbel (- -).

4.5.3 Ajustement d'un modèle global à 4 dimensions

Tel que déjà mentionné, un modèle de copule satisfaisant pour (Co, Cs, Sc, Ti) devra posséder la symétrie radiale à quatre dimensions. Les copules Normal et Student possèdent cette propriété. On pourrait également considérer la version à $d > 2$ dimensions de la copule de Frank, mais ce modèle est peu attrayant car le niveau de dépendance pour chaque paire est identique. Une version à $d > 2$ dimensions de la copule de Plackett a été décrite par Molenberghs et Lesaffre [56], mais comme l'a noté Joe [45], il est ardu de simuler à partir de ce modèle et sa construction implique un nombre élevé de contraintes difficiles à vérifier. Par conséquent, on testera les hypothèses nulles d'une copule Normale et Student à $\nu = 1$, $\nu = 5$ et $\nu = 9$ degrés de liberté. On considérera également la copule Khi-deux centrée, ainsi que les copules de Fisher à $\nu = 1$ et $\nu = 5$ degrés de liberté.

TABLE 4.2 – Résultats du test d'adéquation basé sur $T_{n,N}^{\lambda=3}$ lorsque $N = 250$ pour le vecteur (Co, Cs, Sc, Ti) du jeu de données de Cook et Johnson [12].

Modèle de copule	Estimation de la P-valeur
N	0.327
χ^2	0.034
T_1	0.200
T_5	0.444
T_9	0.883
F_1	0.002
F_5	0.002

Les résultats concernant l'estimation de la P-valeur (basée sur $M = 1\,000$ échantillons *bootstrap*) du test basé sur $T_{n,N}^{\lambda=3}$ sont rapportés dans le Tableau 4.2. En totale cohérence avec les résultats observés dans le cas bidimensionnel, les copules Normale et Student sont acceptés au seuil $\alpha = 0,05$, alors que les copules Khi-deux et Fisher sont rejetés. Il ressort donc clairement de cette analyse que la copule de Student est un bon modèle de dépendance pour le quadruplet (Co, Cs, Sc, Ti).

On note enfin que la valeur de l'estimateur à deux niveaux de type *pairwise* proposé par Quessy et Durocher [60] fournit un degré de liberté estimé à $\hat{\nu} = 6$ sous une copule de Student. On conclut qu'un modèle de copule adéquat pour les données de Cook et Johnson [12] est la Student à 6 degrés de liberté. Pour se convaincre de la justesse de cette découverte et sur sa reproductibilité, on a simulé un jeu de données de taille $n = 655$ à partir de ce modèle ; les nuages de points de ces données simulées sont à la Figure 4.5. Les similitudes sont frappantes entre ces nuages et ceux des vraies données offerts à la Figure 4.3.

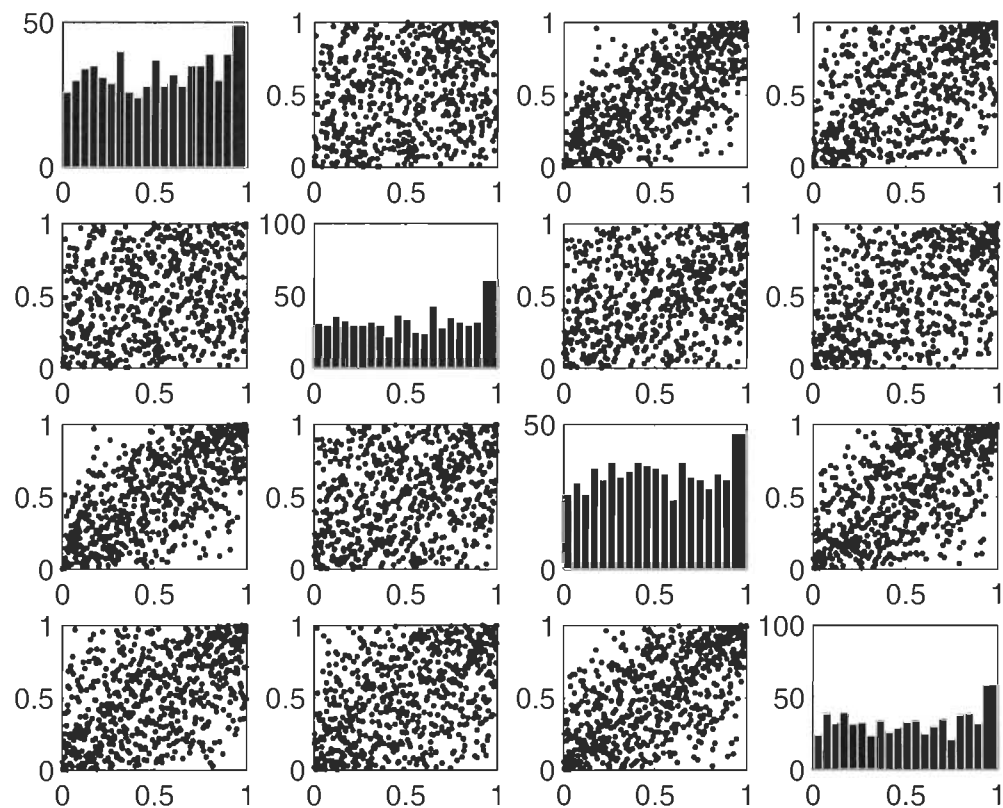


FIGURE 4.5 – Nuages de points (données brutes et rangs standardisés) et histogrammes de $n = 655$ observations simulées de la copule de Student à $\nu = 6$ degrés de liberté et de matrice de corrélation $\hat{\Sigma}$ basée sur l'inversion du tau de Kendall à partir des données de Cook et Johnson [12].

TABLE 4.3 – Estimations, à l'aide de 1 000 réplicats, des seuils de signification des tests d'adéquation basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$.

C	$\tau(C)$	n	$S_{n,N}^{\lambda=1/2}$	$S_{n,N}^{\lambda=1}$	$S_{n,N}^{\lambda=3}$	$S_{n,N}^{\lambda=5}$	$T_{n,N}^{\lambda=1/2}$	$T_{n,N}^{\lambda=1}$	$T_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=5}$	$V_{n,N}^K$	$V_{n,N}^C$
N	.25	100	5.6	4.8	5.0	5.7	5.5	4.8	7.1	3.9	4.3	4.9
		250	4.2	5.7	4.6	4.7	4.5	5.4	4.9	6.7	4.1	4.2
	.50	100	6.5	4.9	4.9	5.3	5.1	5.9	5.2	4.8	5.1	5.3
		250	3.9	5.3	5.0	5.7	6.1	5.6	5.6	5.7	3.8	4.1
	.75	100	2.2	5.0	3.2	4.0	4.7	4.2	4.1	3.3	2.3	3.9
		250	3.4	2.7	4.3	2.5	3.1	4.6	4.2	3.6	3.0	4.0
		100	6.5	5.8	5.3	4.8	4.9	3.9	4.2	4.7	4.0	4.3
		250	4.3	4.0	4.5	4.4	5.6	5.8	3.9	5.1	5.4	5.7
Cℓ	.50	100	6.4	7.2	5.7	6.4	5.0	4.0	4.0	4.6	4.9	3.8
		250	6.3	5.4	5.8	5.3	5.7	4.2	3.6	3.9	5.3	6.6
	.75	100	2.2	3.2	3.2	4.0	3.6	7.6	6.2	6.2	1.5	5.5
		250	4.4	3.8	4.8	3.2	4.6	4.2	6.0	4.8	1.6	3.6
		100	4.6	4.7	4.6	4.9	5.8	6.1	6.3	5.9	5.5	6.2
		250	4.8	3.9	4.2	5.0	5.7	5.9	5.3	5.5	5.7	5.6
	.50	100	5.4	4.6	1.0	6.6	5.7	6.1	5.3	4.0	5.8	6.2
		250	6.5	7.0	5.5	5.5	6.1	4.5	4.5	5.9	4.4	4.2
Gu	.75	100	5.0	2.4	2.8	3.1	3.2	5.6	2.9	4.2	1.1	3.0
		250	2.5	4.5	2.6	4.0	3.4	3.8	4.7	4.3	1.6	5.7

TABLE 4.4 – Estimations, à l'aide de 1 000 répliquats, des pourcentages de rejet de l'hypothèse nulle d'une copule normale pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$.

C	$\tau(C)$	n	$S_{n,N}^{\lambda=1/2}$	$S_{n,N}^{\lambda=1}$	$S_{n,N}^{\lambda=3}$	$S_{n,N}^{\lambda=5}$	$T_{n,N}^{\lambda=1/2}$	$T_{n,N}^{\lambda=1}$	$T_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=5}$	$V_{n,N}^K$	$V_{n,N}^C$
$C\ell$.25	100	5.5	6.2	28.0	31.6	41.9	43.5	45.7	43.9	20.2	27.3
		250	5.8	6.6	27.0	36.5	52.3	47.3	49.8	54.0	22.3	30.8
	.50	100	4.7	15.6	72.4	66.8	83.7	86.7	86.3	85.9	53.5	54.6
		250	6.6	11.7	85.2	84.0	95.6	96.0	96.3	97.6	69.6	76.4
	.75	100	4.9	51.1	93.7	83.7	97.7	97.5	98.5	98.6	55.9	71.9
		250	7.9	40.6	99.1	96.1	99.7	100.0	100.0	99.9	85.1	93.7
Gu	.25	100	4.7	4.6	8.8	7.1	7.7	8.5	8.8	7.4	7.4	7.2
		250	4.5	5.6	13.8	15.6	23.3	23.6	23.8	26.4	9.4	11.3
	.50	100	4.9	4.1	16.4	15.3	19.7	18.1	19.1	18.3	12.7	12.0
		250	6.8	5.5	27.1	26.5	44.3	48.3	50.4	49.8	17.0	19.7
	.75	100	3.4	5.9	13.1	11.0	15.4	16.7	16.3	17.7	6.7	11.3
		250	5.0	5.2	33.8	24.7	47.5	59.0	62.3	54.7	11.8	19.0
$P\ell$.25	100	6.6	4.1	6.2	5.0	4.1	3.7	4.3	4.3	6.6	5.4
		250	4.8	5.1	5.0	7.3	6.0	4.8	4.4	6.2	5.3	5.8
	.50	100	5.0	5.7	7.6	7.2	7.5	5.3	5.6	5.2	8.9	5.2
		250	4.9	5.8	6.1	6.9	7.8	4.2	5.4	4.6	7.7	7.0
	.75	100	2.7	5.4	7.6	6.4	6.8	7.7	6.7	6.8	11.4	6.7
		250	5.1	3.9	7.6	7.1	5.5	9.0	6.4	4.6	10.5	7.2
Fr	.25	100	5.0	5.7	6.0	5.5	4.9	5.7	6.0	4.0	5.5	6.6
		250	4.7	5.5	6.4	8.6	6.4	4.0	4.9	6.6	7.9	6.3
	.50	100	4.7	4.9	9.2	11.2	6.4	4.8	5.1	5.5	10.7	11.9
		250	4.6	4.4	10.3	17.2	11.5	6.6	6.3	5.7	13.9	16.7
	.75	100	1.8	5.2	17.6	21.4	15.7	8.9	7.5	7.9	9.8	21.7
		250	3.0	4.7	20.2	33.7	30.5	14.3	10.9	7.1	23.4	33.6

TABLE 4.5 – Estimations, à l'aide de 1 000 réplicats, des pourcentages de rejet de l'hypothèse nulle d'une copule de Clayton pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$.

C	$\tau(C)$	n	$S_{n,N}^{\lambda=1/2}$	$S_{n,N}^{\lambda=1}$	$S_{n,N}^{\lambda=3}$	$S_{n,N}^{\lambda=5}$	$T_{n,N}^{\lambda=1/2}$	$T_{n,N}^{\lambda=1}$	$T_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=5}$	$V_{n,N}^K$	$V_{n,N}^C$
N	.25	100	5.0	5.9	20.3	19.8	24.2	22.9	21.2	23.1	16.4	14.4
		250	4.4	6.4	24.7	32.3	51.0	53.5	50.9	51.6	26.9	27.5
	.50	100	5.8	14.7	57.9	56.9	69.4	71.0	70.6	72.0	42.0	44.4
		250	5.7	10.6	82.9	83.4	97.4	96.7	97.6	97.3	69.3	75.8
	.75	100	1.6	7.8	78.4	76.1	93.2	95.3	95.6	96.3	32.6	68.7
		250	5.2	16.6	98.4	97.3	99.8	100.0	100.0	99.9	69.5	91.9
		100	6.0	8.8	44.6	45.3	55.3	52.0	51.8	51.9	36.6	36.1
		250	5.3	8.4	63.9	70.0	88.2	89.3	89.1	89.4	59.7	64.5
	Gu	100	8.0	36.8	91.0	89.1	96.6	97.0	96.9	96.9	84.2	83.4
		250	9.7	31.9	99.4	99.7	99.9	100.0	100.0	100.0	96.5	98.9
		100	2.5	36.5	97.7	96.9	99.8	100.0	99.9	100.0	70.6	94.8
		250	9.9	69.0	100.0	100.0	100.0	100.0	100.0	100.0	96.8	100.0
Pl	.25	100	6.1	7.3	19.3	20.9	27.8	22.9	23.0	24.0	20.0	18.4
		250	4.6	6.5	29.0	37.7	60.6	60.6	58.5	59.4	29.1	36.4
	.50	100	5.6	14.4	62.1	60.7	73.2	75.1	72.9	74.3	51.3	48.6
		250	7.0	10.6	81.6	84.6	97.8	98.2	98.0	97.5	67.8	78.9
	.75	100	1.2	12.4	73.3	70.0	89.1	93.9	93.8	94.3	37.7	64.4
		250	6.1	17.9	96.1	94.9	99.7	100.0	100.0	100.0	68.7	89.8
		100	4.5	8.5	21.6	22.6	28.5	25.6	23.7	24.8	18.4	20.9
		250	4.2	5.6	25.8	36.7	55.4	56.6	55.3	55.7	31.2	36.5
	Fr	100	6.7	13.6	58.9	61.2	74.5	71.9	71.5	71.9	49.1	54.6
		250	7.4	9.6	83.2	86.3	98.1	98.1	98.5	98.6	70.4	84.1
		100	1.6	11.7	73.4	71.5	90.1	94.0	93.8	92.5	25.8	67.8
		250	6.0	17.3	96.3	96.7	99.9	100.0	100.0	100.0	60.4	92.2

TABLE 4.6 – Estimations, à l'aide de 1 000 réplicats, des pourcentages de rejet de l'hypothèse nulle d'une copule de Gumbel pour les tests basés sur $S_{n,N}^{\lambda=1/2}$, $S_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $S_{n,N}^{\lambda=5}$, $T_{n,N}^{\lambda=1/2}$, $T_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=5}$, $V_{n,N}^K$ et $V_{n,N}^C$.

C	$\tau(C)$	n	$S_{n,N}^{\lambda=1/2}$	$S_{n,N}^{\lambda=1}$	$S_{n,N}^{\lambda=3}$	$S_{n,N}^{\lambda=5}$	$T_{n,N}^{\lambda=1/2}$	$T_{n,N}^{\lambda=1}$	$T_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=5}$	$V_{n,N}^K$	$V_{n,N}^C$
\mathcal{Cl}	.25	100	4.2	6.3	14.8	13.6	19.1	22.4	22.0	20.2	10.2	14.9
		250	5.2	5.2	14.1	14.8	23.3	22.5	24.1	24.7	12.9	12.0
	.50	100	6.0	7.6	25.6	24.1	34.2	36.6	35.5	34.7	14.4	16.9
		250	7.5	8.2	26.1	27.2	42.3	43.1	45.1	45.4	19.2	22.8
	.75	100	3.6	2.6	20.1	21.0	33.4	36.1	35.2	40.8	2.7	8.9
		250	3.6	6.1	27.1	24.4	42.8	47.6	49.6	50.5	8.4	18.9
\mathcal{N}	.25	100	5.3	10.1	56.6	55.7	72.6	75.6	72.3	73.8	43.7	53.0
		250	5.4	10.0	66.7	73.9	87.6	87.2	87.7	87.6	58.5	65.2
	.50	100	10.3	42.8	95.4	92.8	98.8	98.8	99.0	98.8	85.1	88.6
		250	7.3	38.9	99.8	98.5	100.0	100.0	99.8	100.0	97.6	99.4
	.75	100	9.6	59.9	99.4	98.7	100.0	100.0	99.9	100.0	68.9	93.8
		250	6.4	80.3	100.0	100.0	100.0	100.0	100.0	100.0	97.9	99.9
\mathcal{Pl}	.25	100	4.7	7.0	15.6	12.9	18.8	20.5	20.1	19.7	11.1	13.0
		250	6.0	5.3	12.9	16.4	20.1	20.7	20.2	19.7	13.5	12.1
	.50	100	6.4	8.5	26.1	22.1	32.3	32.2	30.8	31.8	17.9	17.3
		250	4.6	6.5	25.2	29.3	36.8	35.3	37.4	41.4	24.3	23.0
	.75	100	4.0	2.7	23.4	20.6	33.2	37.7	33.8	40.0	5.6	14.0
		250	3.2	4.8	25.7	27.0	40.6	44.1	42.9	42.0	11.0	23.4
\mathcal{Fr}	.25	100	3.9	5.8	14.4	15.5	18.8	18.6	19.7	18.5	11.3	12.7
		250	5.6	5.1	14.1	18.8	18.6	21.2	20.2	19.8	14.3	13.6
	.50	100	7.7	9.2	32.8	30.6	31.8	34.2	31.4	31.8	21.4	25.3
		250	4.7	7.3	33.6	43.4	43.7	41.0	41.0	40.5	29.2	36.2
	.75	100	5.3	6.3	33.0	41.1	43.8	40.8	34.2	41.0	5.8	29.8
		250	3.3	5.1	42.5	57.2	60.6	47.4	46.6	45.1	19.2	53.4

Conclusion

Le point de départ de ce travail de recherche est l'exploitation de la fonction caractéristique associée à la variable aléatoire intégrale de probabilité multidimensionnelle W pour tester les hypothèses \mathbb{H}_0 et \mathbb{H}_1 d'adéquation à une famille de copules. On a ainsi défini les fonctions caractéristique et caractéristique pondérée de W , à savoir

$$\psi_W(t) = \mathbb{E}(e^{itW}) \quad \text{et} \quad \psi_W^\lambda(t) = \mathbb{E}\left\{(W(1-W))^{\lambda|t|} e^{itW}\right\}.$$

Les statistiques de test qui ont été proposées pour confronter \mathbb{H}_0 à \mathbb{H}_1 sont

$$S_{n,N}^g = n \int_{\mathbb{R}} |\psi_n(t) - \psi_N(t)|^2 g(t) dt \quad \text{et} \quad T_{n,N}^\lambda = n \int_{\mathbb{R}} |\psi_n^\lambda(t) - \psi_N^\lambda(t)|^2 dt.$$

Les simulations ont montré clairement que les tests basés sur ces statistiques sont des alternatives avantageuses aux tests d'adéquation populaires. En effet, ces tests conservent leur seuil nominal sous \mathbb{H}_0 et sont généralement plus puissants que leurs concurrents. Dans des travaux futurs, il serait intéressant d'explorer davantage le cas multidimensionnel à l'aide de simulations. Effectivement, dans le cadre de ce travail, seul le cas bivarié a été étudié par simulations. Une telle étude dans le cas d -dimensionnel serait donc intéressante pour quiconque souhaite effectuer de la modélisation à plusieurs variables. Enfin, à la lumière des résultats des Figures 3.3 et 3.4, on pourrait considérer des tests basés uniquement sur la partie imaginaire de ψ_W^λ , car sa partie réelle semble peu apte à distinguer entre les différents modèles de copules.

Bibliographie

- [1] V. ALBA FERNÁNDEZ, M. D. JIMÉNEZ GAMERO et J. MUNOZ GARCIA : A test for the two-sample problem based on empirical characteristic functions. *Comput. Statist. Data Anal.*, 52(7):3730–3748, 2008.
- [2] T. BAHRAOUI, T. BOUEZMARNI et J.-F. QUESSY : A family of goodness-of-fit tests for copulas based on characteristic functions. *Scand. J. Stat.*, 45(2):301–323, 2018.
- [3] T. BAHRAOUI, T. BOUEZMARNI et J.-F. QUESSY : Testing the symmetry of a dependence structure with a characteristic function. *Depend. Model.*, 6(1):331–355, 2018.
- [4] T. BAHRAOUI et J.-F. QUESSY : Tests of radial symmetry for multivariate copulas based on the copula characteristic function. *Electron. J. Stat.*, 2017.
- [5] P. BARBE, C. GENEST, K. GHOUDI et B. RÉMILLARD : On Kendall’s process. *J. Multivariate Anal.*, 58(2):197–229, 1996.
- [6] L. BARINGHAUS et N. HENZE : A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, 35(6):339–348, 1988.
- [7] D. BERG : Copula goodness-of-fit testing : an overview and power comparison. *The European Journal of Finance*, 15(7-8):675–701, 2009.
- [8] A. BÀRDOSSY : Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, 42(11):1–12, 2006.

- [9] S. CAMBANIS, S. HUANG et G. SIMONS : On the theory of elliptically contoured distributions. *J. Multivariate Anal.*, 11(3):368–385, 1981.
- [10] G. CASELLA et R. L. BERGER : *Statistical inference*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.
- [11] D. G. CLAYTON : A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151, 1978.
- [12] R. D. COOK et M. E. JOHNSON : Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. *Technometrics*, 28(2):123–131, 1986.
- [13] S. CSORGŐ : Limit behaviour of the empirical characteristic function. *Ann. Probab.*, 9(1):130–144, 1981.
- [14] S. CSORGŐ : Multivariate empirical characteristic functions. *Z. Wahrsch. Verw. Gebiete*, 55(2):203–229, 1981.
- [15] G. DALL'AGLIO : Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Scuola Norm. Sup. Pisa (3)*, 10:35–74, 1956.
- [16] P. DEHEUVELS : La fonction de dépendance empirique et ses propriétés. un test non paramétrique d'indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, 65:274–292, 1979.
- [17] P. DEHEUVELS : Nonparametric test of independence. In *Nonparametric asymptotic statistics (Proc. Conf., Rouen, 1979) (French)*, vol. 821 de *Lecture Notes in Math.*, p. 95–107. Springer, Berlin, 1980.
- [18] P. DEHEUVELS : Multivariate tests of independence. In *Analytical methods in probability theory (Oberwolfach, 1980)*, vol. 861 de *Lecture Notes in Math.*, p. 42–50. Springer, Berlin-New York, 1981.

- [19] S. DEMARTA et A. J. MCNEIL : The T Copula and Related Copulas. *International Statistical Review*, 73(1):111–129, avr. 2005.
- [20] T. W. EPPS : Tests for location-scale families based on the empirical characteristic function. *Metrika*, 62(1):99–114, 2005.
- [21] T. W. EPPS et L. B. PULLEY : A test for normality based on the empirical characteristic function. *Biometrika*, 70(3):723–726, 1983.
- [22] H.-B. FANG, K.-T. FANG et S. KOTZ : The meta-elliptical distributions with given marginals. *J. Multivariate Anal.*, 82(1):1–16, 2002.
- [23] A.-C. FAVRE, J.-F. QUESSY et M.-H. TOUPIN : The new family of Fischer copulas to model upper tail dependence and radial asymmetry : properties and application to high-dimensional rainfall data. *Environmetrics*, 29(3), 2018.
- [24] R. FÉRON : Sur les tableaux de corrélation dont les marges sont données. Cas de l'espace a trois dimensions. *Publ. Inst. Statist. Univ. Paris*, 5:3–12, 1956.
- [25] A. FEUERVERGER et R. A. MUREIKA : The empirical characteristic function and its applications. *Ann. Statist.*, 5(1):88–97, 1977.
- [26] K. FRAGIADAKIS et S. G. MEINTANIS : Goodness-of-fit tests for multivariate Laplace distributions. *Math. Comput. Modelling*, 53(5-6):769–779, 2011.
- [27] M. FRÉCHET : Sur les tableaux de corrélation dont les marges sont données. *Ann. Univ. Lyon. Sect. A. (3)*, 14:53–77, 1951.
- [28] C. GENEST, K. GHOUDI et L.-P. RIVEST : A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82(3):543–552, 1995.
- [29] C. GENEST, I. KOJADINOVIC, J. NEŠLEHOVÁ et J. YAN : A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, 16, 2010.
- [30] C. GENEST : Frank's family of bivariate distributions. *Biometrika*, 74(3):549–555, 1987.

- [31] C. GENEST, W. HUANG et J.-M. DUFOUR : A regularized goodness-of-fit test for copulas. *J. SFdS*, 154(1):64–77, 2013.
- [32] C. GENEST et J. MACKAY : The joy of copulas : bivariate distributions with uniform marginals. *Amer. Statist.*, 40(4):280–283, 1986.
- [33] C. GENEST, J.-F. QUESSY et B. RÉMILLARD : Goodness-of-fit procedures for copula models based on the probability integral transformation. *Scand. J. Statist.*, 33(2):337–366, 2006.
- [34] C. GENEST et L.-P. RIVEST : A characterization of Gumbel’s family of extreme value distributions. *Statist. Probab. Lett.*, 8(3):207–211, 1989.
- [35] C. GENEST et L.-P. RIVEST : Statistical inference procedures for bivariate archimedean copulas. *J. Amer. Statist. Assoc.*, 88(423):1034–1043, 1993.
- [36] C. GENEST et B. RÉMILLARD : Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(6):1096–1127, 2008.
- [37] C. GENEST, B. RÉMILLARD et D. BEAUDOIN : Goodness-of-fit tests for copulas : a review and a power study. *Insurance Math. Econom.*, 44(2):199–213, 2009.
- [38] E. J. GUMBEL : Bivariate logistic distributions. *J. Amer. Statist. Assoc.*, 56:335–349, 1961.
- [39] N. GÜRTLER et N. HENZE : Goodness-of-fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, 52(2):267–286, 2000.
- [40] N. HENZE, B. KLAR et S. G. MEINTANIS : Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *J. Multivariate Anal.*, 87(2):275–297, 2003.
- [41] W. HÖFFDING : Maszstabinvariante Korrelationstheorie. *Schr. Math. Inst. u. Inst. Angew. Math. Univ. Berlin*, 5:181–233, 1940.

- [42] P. HOUGAARD : A class of multivariate failure time distributions. *Biometrika*, 73(3):671–678, 1986.
- [43] M. D. JIMÉNEZ-GAMERO et H.-M. KIM : Fast goodness-of-fit tests based on the characteristic function. *Comput. Statist. Data Anal.*, 89:172–191, 2015.
- [44] M. JIMÉNEZ-GAMERO, V. ALBA-FERNÁNDEZ, J. MUNOZ-GARCIA et Y. CHALCO-CANO : Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics & Data Analysis*, 53(12):3957–3971, 2009.
- [45] H. JOE : *Dependence modeling with copulas*, vol. 134 de *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2015.
- [46] M. G. KENDALL : A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.
- [47] H.-M. KIM et M. G. GENTON : Characteristic functions of scale mixtures of multivariate skew-normal distributions. *J. Multivariate Anal.*, 102(7):1105–1117, 2011.
- [48] G. KIMELDORF et A. SAMPSON : Uniform representations of bivariate distributions. *Comm. Statist.*, 4(7):617–627, 1975.
- [49] B. KLAR et S. G. MEINTANIS : Tests for normal mixtures based on the empirical characteristic function. *Comput. Statist. Data Anal.*, 49(1):227–242, 2005.
- [50] T. J. KOZUBOWSKI, K. PODGÓRSKI et I. RYCHLIK : Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*, 113:59–72, 2013.
- [51] M. MATSUI et A. TAKEMURA : Empirical characteristic function approach to goodness-of-fit tests for the Cauchy distribution with parameters estimated by mle or eise. *Ann. Inst. Statist. Math.*, 57(1):183–199, 2005.

- [52] S. MEINTANIS, J. SWANEPOEL et J. ALLISON : The probability weighted characteristic function and goodness-of-fit testing. *Journal of statistical planning and inference*, 146:122–132, 2014.
- [53] S. G. MEINTANIS et G. ILIOPOULOS : Fourier methods for testing multivariate independence. *Comput. Statist. Data Anal.*, 52(4):1884–1895, 2008.
- [54] S. G. MEINTANIS et N. G. USHAKOV : Nonparametric probability weighted empirical characteristic function and applications. *Statist. Probab. Lett.*, 108:52–61, 2016.
- [55] M. MESFIOUI, J.-F. QUESSY et M.-H. TOUPIN : On a new goodness-of-fit process for families of copulas. *Canad. J. Statist.*, 37(1):80–101, 2009.
- [56] G. MOLENBERGHS et E. LESAFFRE : Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association*, 89(426):633–644, 1994.
- [57] M. OMELKA, I. GIJBELS et N. VERAVERBEKE : Improved kernel estimation of copulas : weak convergence and goodness-of-fit testing. *Ann. Statist.*, 37(5B):3023–3058, 2009.
- [58] R. L. PLACKETT : A class of bivariate distributions. *J. Amer. Statist. Assoc.*, 60:516–522, 1965.
- [59] J.-F. QUESSY : *Théorie et application des copules : tests d'adéquation, tests d'indépendance, et bornes pour la valeur-à-risque*. Thèse de doctorat, Université Laval, Québec, Canada, 2005.
- [60] J.-F. QUESSY et M. DUROCHER : The class of copulas arising from squared distributions : properties and inference. *Econometrics and Statistics*, 2019.
- [61] J.-F. QUESSY, L.-P. RIVEST et M.-H. TOUPIN : On the family of multivariate chi-square copulas. *J. Multivariate Anal.*, 152:40–60, 2016.

- [62] O. SCAILLET : Kernel-based goodness-of-fit tests for copulas with fixed smoothing parameters. *J. Multivariate Anal.*, 98(3):533–543, 2007.
- [63] J. SHAO : *Mathematical statistics*. Springer Texts in Statistics. Springer-Verlag, New York, second édn, 2003.
- [64] J. H. SHIH et T. A. LOUIS : Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, 51(4):1384–1399, 1995.
- [65] A. SKLAR : Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, 8:229–231, 1959.
- [66] W. STUTE, W. GONZÁLEZ-MANTEIGA et M. PRESEDO-QUINDIMIL : Bootstrap based goodness-of-fit tests. *Metrika*, 40:243–256, 1993.

Annexe A

Éléments complémentaires

A.1 Quelques démonstrations

A.1.1 Symétrie radiale pour la copule de Frank

De la définition d'une copule de survie présentée à la Section 2.3.3, on obtient

$$\widehat{C}_\theta(u_1, u_2) = u_1 + u_2 - 1 - \frac{1}{\theta} \ln \left\{ \frac{-e^{-\theta} + e^{-\theta(1-u_1)} + e^{-\theta(1-u_2)} - e^{-\theta(2-u_1-u_2)}}{1 - e^{-\theta}} \right\}.$$

En appliquant l'exponentielle de chaque côté de l'expression précédente, on trouve

$$\begin{aligned} \exp \left\{ \widehat{C}_\theta(u_1, u_2) \right\} &= e^{u_1+u_2-1} \left(\frac{-e^{-\theta} + e^{-\theta(1-u_1)} + e^{-\theta(1-u_2)} - e^{-\theta(2-u_1-u_2)}}{1 - e^{-\theta}} \right)^{-1/\theta} \\ &= \left\{ \frac{-e^{-\theta(u_1+u_2)} + e^{-\theta u_2} + e^{-\theta u_1} - e^{-\theta}}{1 - e^{-\theta}} \right\}^{-1/\theta} \\ &= \exp \{ C_\theta(u_1, u_2) \}. \end{aligned}$$

On vient donc de montrer que $\widehat{C}_\theta = C_\theta$, concluant donc la preuve.

A.1.2 Dérivation de K_W pour les copules Archimédiennes bi-dimensionnelles

En premier lieu, on note que

$$C_v^{-1}(t) = \phi^{-1} \{ \phi(t) - \phi(v) \} \quad \text{et} \quad C^{(01)}(u_1, u_2) = \frac{\phi'(u_2)}{\phi' \circ \phi^{-1} \{ \phi(u_1) + \phi(u_2) \}}.$$

Ainsi, puisque $\phi(1) = 0$,

$$\begin{aligned} K_W(w) &= w + \int_w^1 \frac{\phi'(v)}{\phi' \circ \phi^{-1} \{ \phi(w) - \phi(v) + \phi(v) \}} dv \\ &= w + \frac{1}{\phi'(w)} \int_w^1 \phi'(v) dv \\ &= w + \frac{1}{\phi'(w)} \{ \phi(1) - \phi(w) \} \\ &= w - \frac{\phi(w)}{\phi'(w)}. \end{aligned}$$

A.2 Estimation du paramètre d'une copule

A.2.1 Rappel sur les méthodes classiques d'estimation

Supposons que l'on dispose d'observations indépendantes X_1, \dots, X_n tirées d'une certaine population. En général, on fait l'hypothèse que la loi sous-jacente de cette population appartient à une famille de lois de probabilité F_θ , où $\theta \in \Theta \subseteq \mathbb{R}^p$. Pour estimer θ dans ce contexte, deux approches classiques seront présentées : la méthode des moments et la méthode du maximum de vraisemblance.

Considérons le cas à un paramètre, c'est-à-dire $\theta \in \mathbb{R}$. La méthode des moments consiste à poser $m_1(\theta) = E(X)$, où $X \sim F_\theta$. On définit alors l'estimateur des moments $\hat{\theta}^{\text{MM}}$ de θ comme la solution de

$$m_1(\theta) = \bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j.$$

Lorsque m_1 est continue et inversible, on a $\hat{\theta}^{\text{MM}} = m_1^{-1}(\bar{X})$.

Dans un autre ordre d'idée, considérons la méthode du maximum de vraisemblance. Celle-ci est basée sur la fonction de log-vraisemblance de θ définie par

$$L(\theta) = \sum_{j=1}^n \ln f_\theta(X_j),$$

où f_θ est la densité de F_θ . L'estimateur à maximum de vraisemblance de θ est la valeur $\hat{\theta}^{\text{MV}} \in \Theta$ qui maximise $L(\theta)$, à savoir

$$\hat{\theta}^{\text{MV}} = \arg \max_{\theta \in \Theta} L(\theta).$$

En général, si $\theta = (\theta_1, \dots, \theta_K) \in \mathbb{R}^K$ et que toutes les dérivées partielles existent, le vecteur $\hat{\theta}^{\text{MV}} = (\hat{\theta}_1^{\text{MV}}, \dots, \hat{\theta}_K^{\text{MV}})$ est une solution possible, tel que

$$\left. \frac{\partial}{\partial \theta_j} \tilde{L}(\theta) \right|_{\theta = \hat{\theta}^{\text{MV}}} = 0, \quad j \in \{1, \dots, K\}.$$

A.2.2 Adaptation aux copules

Une adaptation des méthodes, dites classiques, pour nos modèles de dépendance est requise. Comme les méthodes des moments et de vraisemblance ne peuvent s'appliquer directement, les méthodes d'inversion du tau de Kendall et par maximum de pseudo-vraisemblance seront plutôt utilisées.

Considérons d'abord une famille de copules $\{C_\theta; \theta \in \Theta\}$, où $\Theta \subseteq \mathbb{R}$. Dans ce cas, selon la définition donnée à l'Équation (2.3), le tau de Kendall associé à C_θ s'écrit

$$\tau = \kappa(\theta) = 4 \int_0^1 \int_0^1 C_\theta(u, v) dC_\theta(u, v) du dv - 1.$$

En s'inspirant de la méthode des moments classiques, une idée consiste à estimer θ avec la valeur θ_n telle que $\kappa(\theta_n) = \tau_n$, où τ_n est le tau de Kendall empirique, c'est-à-dire

$$\tau_n = \frac{4}{n(n-1)} \sum_{j < k} \mathbf{1}\{(X_j - X_k)(Y_j - Y_k) > 0\} - 1.$$

Ainsi, en autant que la fonction κ est inversible, l'estimateur de θ par inversion du tau de Kendall est défini par

$$\theta_n^{\text{MM}} = \kappa^{-1}(\tau_n).$$

Dans le cas où κ n'admet pas d'inverse explicite, il faut solutionner numériquement l'équation $\kappa(\theta_n) = \tau_n$. À noter que cette méthode peut être utilisée seulement lorsque le paramètre θ est univarié. Si $\theta = (\theta_1, \dots, \theta_K) \in \Theta \subseteq \mathbb{R}^K$, il faudra plutôt recourir à une méthode basée sur la vraisemblance.

Exemple A.1. *Par une simple application de la définition, le tau de Kendall pour la copule paramétrique de Farlie–Gumbel–Morgenstern est donné par $\kappa(\theta) = 2\theta/9$. On trouve alors de manière directe $\kappa^{-1}(t) = 9t/2$, et donc $\theta_n^{\text{MM}} = 9\tau_n/2$.*

Dans le même ordre d'idée, il est possible de modifier la méthode par maximum de vraisemblance. En effet, étudiée par Genest *et al.* [28] et par Shih et Louis [64], la méthode par maximum de pseudo-vraisemblance est similaire à celle de l'IFM présentée précédemment. De manière générale, au lieu de remplacer les marges inconnues par des modèles paramétriques, il suffit plutôt d'estimer F et G par les fonctions de répartition empiriques définies par

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(X_i \leq x) \quad \text{et} \quad G_n(y) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{1}(Y_i \leq y).$$

Bien entendu, il faut que la densité c_θ de C_θ existe. De plus, il est possible d'estimer les paramètres pour $\theta \in \mathbb{R}^k$. Considérons R_i le rang de X_i parmi X_1, \dots, X_n et S_i le rang de Y_i parmi Y_1, \dots, Y_n . Donc, naturellement on a $F_n(X_i) = R_i/(n+1)$ et $G_n(Y_i) = S_i/(n+1)$ pour tout $i \in \{1, \dots, n\}$. De ce fait, on définit la fonction de pseudo log-vraisemblance par

$$\tilde{L}(\theta) = \sum_{i=1}^n \ell_\theta \{F_n(X_i), G_n(Y_i)\} = \sum_{i=1}^n \ell_\theta \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right),$$

où $\ell_\theta(x, y) = \log C_\theta(x, y)$. À noter qu'on divise par $n+1$ plutôt que par n pour éviter des problèmes numériques. Par conséquent, en utilisant des méthodes numériques d'optimisation, on trouve

$$\hat{\theta}^{\text{MV}} = \arg \max_{\theta \in \Theta} \tilde{L}(\theta).$$

A.3 Détails sur une statistique de test basée sur la copule empirique

A.3.1 Estimation non-paramétrique d'une copule

Soit un échantillon $\mathbf{X}_1, \dots, \mathbf{X}_n$ i.i.d. $F_{\mathbf{X}}$. Une estimation non-paramétrique de $F_{\mathbf{X}}$ s'obtient de la fonction de répartition empirique, à savoir

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}(\mathbf{X}_j \leq \mathbf{x}).$$

Il est possible d'affirmer que F_n est un estimateur sans-biais pour $F_{\mathbf{X}}$. En effet,

$$\mathbb{E} \{F_n(\mathbf{x})\} = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(\mathbf{X}_i \leq \mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}).$$

Pour estimer la copule C sous-jacente à $F_{\mathbf{X}}$, on rappelle d'abord que

$$C(\mathbf{u}) = F_{\mathbf{X}} \{F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)\}.$$

Un estimateur *plug-in* de C tel que proposé par Deheuvels [16] est alors

$$C_n^*(\mathbf{u}) = F_n \{F_{n1}^{-1}(u_1), \dots, F_{nd}^{-1}(u_d)\},$$

où F_{n1}, \dots, F_{nd} sont les fonctions de répartition empiriques marginales et $F_{n1}^{-1}, \dots, F_{nd}^{-1}$ sont leurs inverses généralisés. Habituellement, on utilise plutôt une version asymptotiquement équivalente à C_n^* , à savoir la copule empirique

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{j=1}^n \mathbf{1} \{F_{n1}(X_{j1}) \leq u_1, \dots, F_{nd}(X_{jd}) \leq u_d\}. \quad (\text{A.1})$$

A.3.2 Définition de la statistique de test

Afin de tester l'adéquation à une famille de copules $\{C_\theta; \theta \in \Theta\}$, soit d'abord θ_n , un estimateur convergent de θ obtenu par les méthodes proposées à la Section A.2.2. Une statistique de test basée sur la copule empirique est définie par

$$V_n^C = n \int_0^1 \int_0^1 \{C_n(u, v) - C_{\theta_n}(u, v)\}^2 du dv.$$

Sans surprise, il est extrêmement difficile de trouver une forme explicite pour notre intégrale. Pour remédier à cette situation, on utilisera la version alternative de V_n^C

$$V_{n,N}^C = n \int_0^1 \int_0^1 \{C_n(u, v) - C_N(u, v)\}^2 du dv, \quad (\text{A.2})$$

où C_N est la copule empirique d'observations simulées artificiellement de la loi C_{θ_n} , c'est-à-dire $(X_1^*, Y_1^*), \dots, (X_N^*, Y_N^*)$ i.i.d. C_{θ_n} . Spécifiquement, si R_i^* est le rang de X_i^*

parmi X_1^*, \dots, X_n^* et S_i^* est le rang de Y_i^* parmi Y_1^*, \dots, Y_n^* , alors

$$C_N(u, v) = \frac{1}{N} \sum_{j=1}^N \mathbf{1} \left(\frac{R_j^*}{N} \leq u, \frac{S_j^*}{N} \leq v \right).$$

La statistique de test $V_{n,N}^C$ possède l'avantage de s'écrire de façon explicite. En effet, il est montré à la sous-section suivante que

$$\begin{aligned} V_{n,N}^C = & \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^n \left\{ 1 - \frac{\max(R_j, R_k)}{n} \right\} \left\{ 1 - \frac{\max(S_j, S_k)}{n} \right\} \\ & - \frac{2}{N} \sum_{j=1}^n \sum_{k=1}^N \left\{ 1 - \max \left(\frac{R_j}{n}, \frac{R_k^*}{N} \right) \right\} \left\{ 1 - \max \left(\frac{S_j}{n}, \frac{S_k^*}{N} \right) \right\} \\ & + \frac{n}{N^2} \sum_{j=1}^N \sum_{k=1}^N \left\{ 1 - \frac{\max(R_j^*, R_k^*)}{N} \right\} \left\{ 1 - \frac{\max(S_j^*, S_k^*)}{N} \right\}. \end{aligned} \quad (\text{A.3})$$

Annexe B

Tableaux et figures supplémentaires

TABLE B.1 – Valeurs critiques au seuil $\alpha = 0,05$ pour tester la copule Normale.

$\tau(C)$	n	$S_{n,N}^{\lambda=1/2}$	$S_{n,N}^{\lambda=1}$	$S_{n,N}^{\lambda=3}$	$S_{n,N}^{\lambda=5}$	$T_{n,N}^{\lambda=1/2}$	$T_{n,N}^{\lambda=1}$	$T_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=5}$	$V_{n,N}^K$	$V_{n,N}^C$
.00	100	0,0059	0,0218	0,1808	0,4511	0,1496	0,0707	0,0221	0,0133	0,0519	0,1695
	250	0,0137	0,0525	0,3825	0,8288	0,2317	0,0869	0,0323	0,0191	0,0816	0,2793
.05	100	0,0059	0,0247	0,1770	0,5490	0,1674	0,0757	0,0230	0,0140	0,0520	0,1770
	250	0,0134	0,0495	0,3732	0,7881	0,2461	0,1014	0,0320	0,0193	0,0793	0,2740
.10	100	0,0055	0,0228	0,1738	0,4780	0,1623	0,0746	0,0225	0,0125	0,0506	0,1679
	250	0,0121	0,0467	0,3777	0,8056	0,2336	0,0963	0,0302	0,0191	0,0808	0,2852
.15	100	0,0046	0,0219	0,1937	0,4809	0,1634	0,0719	0,0243	0,0137	0,0510	0,1774
	250	0,0117	0,0438	0,3434	0,7651	0,2115	0,0991	0,0299	0,0184	0,0791	0,2914
.20	100	0,0051	0,0213	0,1965	0,5031	0,1666	0,0672	0,0228	0,0127	0,0496	0,1770
	250	0,0125	0,0441	0,3808	0,8271	0,2454	0,1031	0,0333	0,0195	0,0758	0,2949
.25	100	0,0048	0,0196	0,1568	0,4763	0,1612	0,0639	0,0198	0,0125	0,0442	0,1623
	250	0,0110	0,0418	0,3699	0,8126	0,2338	0,0876	0,0301	0,0173	0,0755	0,2744
.30	100	0,0045	0,0182	0,1742	0,5205	0,1335	0,0594	0,0209	0,0132	0,0424	0,1671
	250	0,0108	0,0425	0,3506	0,7816	0,2064	0,0968	0,0276	0,0164	0,0676	0,2613
.35	100	0,0041	0,0153	0,1605	0,5033	0,1342	0,0633	0,0184	0,0114	0,0417	0,1645
	250	0,0108	0,0352	0,2998	0,7352	0,1979	0,0822	0,0267	0,0151	0,0637	0,2542
.40	100	0,0037	0,0158	0,1598	0,4527	0,1249	0,0568	0,0180	0,0122	0,0391	0,1645
	250	0,0095	0,0361	0,2945	0,7487	0,1856	0,0762	0,0242	0,0138	0,0592	0,2471
.45	100	0,0038	0,0142	0,1502	0,4609	0,1469	0,0565	0,0180	0,0106	0,0357	0,1573
	250	0,0078	0,0343	0,2596	0,7169	0,1854	0,0768	0,0255	0,0128	0,0524	0,2300
.50	100	0,0026	0,0108	0,1424	0,3843	0,1175	0,0478	0,0153	0,0093	0,0315	0,1427
	250	0,0070	0,0266	0,2311	0,6171	0,1644	0,0695	0,0200	0,0137	0,0453	0,2322
.55	100	0,0029	0,0101	0,1143	0,3855	0,1056	0,0446	0,0126	0,0077	0,0281	0,1440
	250	0,0057	0,0228	0,2079	0,5844	0,1538	0,0592	0,0168	0,0105	0,0397	0,2135
.60	100	0,0019	0,0096	0,1061	0,3408	0,0954	0,0400	0,0136	0,0079	0,0240	0,1330
	250	0,0045	0,0199	0,1597	0,5261	0,1241	0,0508	0,0158	0,0097	0,0338	0,1969
.65	100	0,0017	0,0066	0,0865	0,2845	0,0747	0,0327	0,0103	0,0059	0,0206	0,1194
	250	0,0038	0,0154	0,1580	0,4507	0,1050	0,0403	0,0127	0,0086	0,0288	0,1819
.70	100	0,0012	0,0053	0,0690	0,2663	0,0636	0,0268	0,0084	0,0049	0,0159	0,1062
	250	0,0029	0,0132	0,1159	0,3663	0,0854	0,0335	0,0098	0,0061	0,0225	0,1530
.75	100	0,0009	0,0041	0,0514	0,1959	0,0497	0,0193	0,0062	0,0038	0,0133	0,0942
	250	0,0022	0,0081	0,0984	0,2906	0,0658	0,0222	0,0080	0,0044	0,0162	0,1302
.80	100	0,0006	0,0029	0,0371	0,1451	0,0367	0,0158	0,0052	0,0026	0,0094	0,0780
	250	0,0014	0,0056	0,0629	0,1927	0,0434	0,0165	0,0052	0,0029	0,0098	0,1096
.85	100	0,0004	0,0015	0,0219	0,0838	0,0249	0,0104	0,0031	0,0018	0,0065	0,0555
	250	0,0009	0,0036	0,0342	0,1248	0,0266	0,0103	0,0030	0,0018	0,0060	0,0774
.90	100	0,0002	0,0007	0,0119	0,0468	0,0136	0,0063	0,0019	0,0012	0,0042	0,0421
	250	0,0004	0,0015	0,0166	0,0573	0,0117	0,0044	0,0014	0,0008	0,0028	0,0533
.95	100	0,0000	0,0002	0,0051	0,0172	0,0065	0,0030	0,0010	0,0006	0,0025	0,0222
	250	0,0001	0,0005	0,0047	0,0156	0,0036	0,0014	0,0004	0,0003	0,0008	0,0261

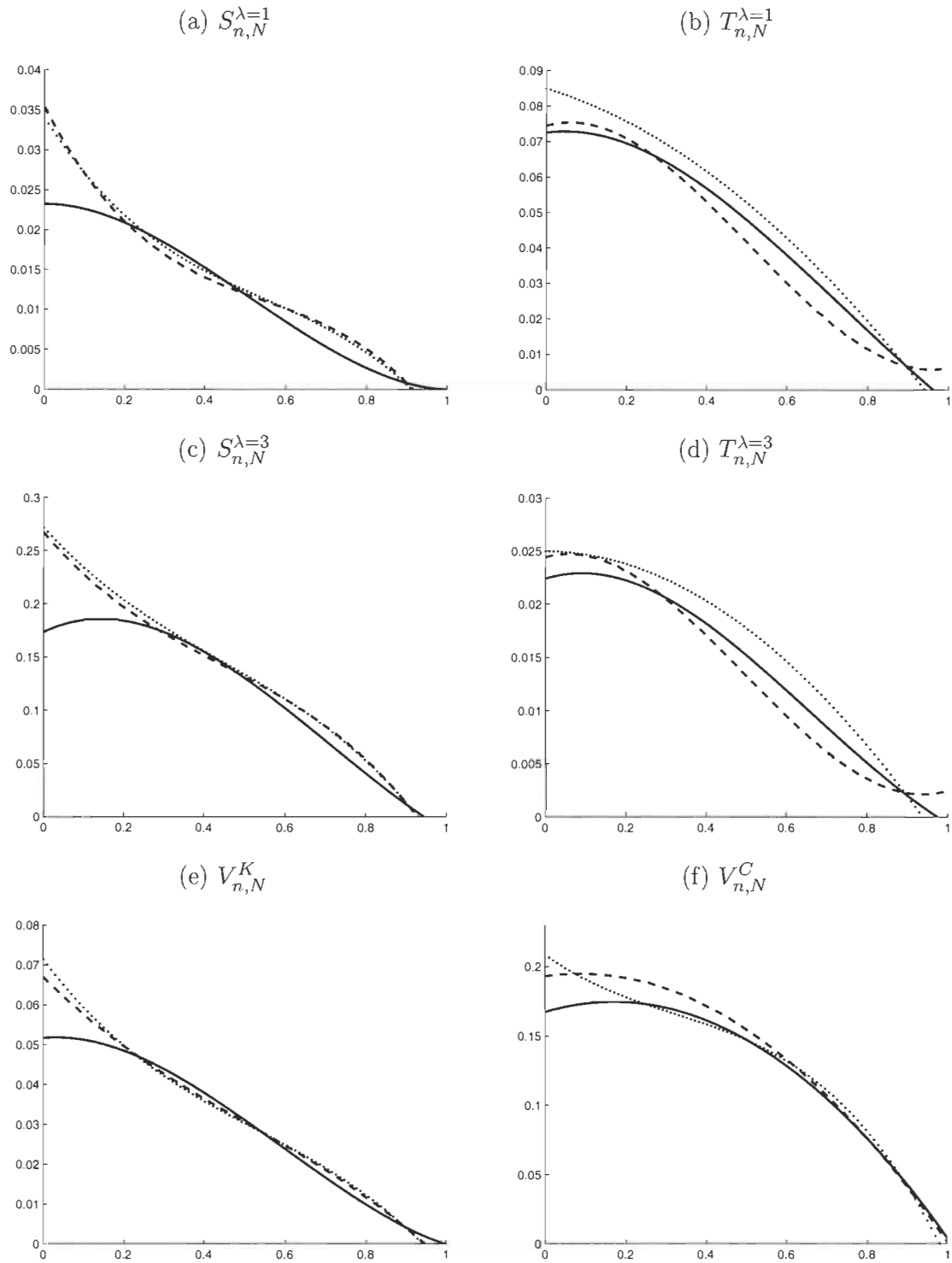


FIGURE B.1 – Régressions polynomiales des valeurs critiques en fonction de τ pour les statistiques de test $S_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ sous les copules Normale (—), Clayton (--) et Gumbel (··) lorsque $n = 100$.

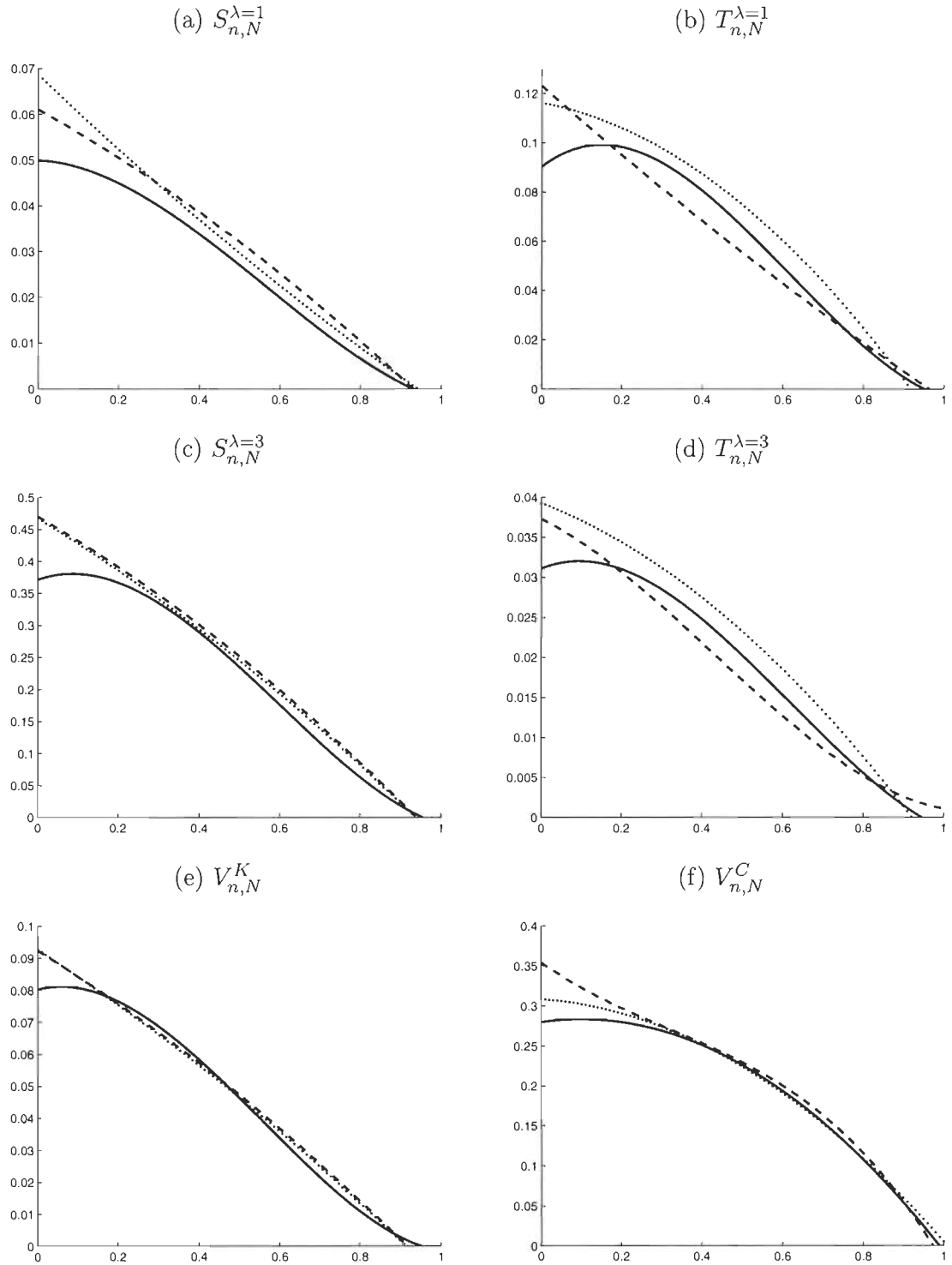


FIGURE B.2 – Régressions polynomiales des valeurs critiques en fonction de τ pour les statistiques de test $S_{n,N}^{\lambda=1}$, $T_{n,N}^{\lambda=1}$, $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ sous les copules Normale (—), Clayton (--) et Gumbel (· ·) lorsque $n = 250$.

TABLE B.2 – P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Co,Cs) et (Co,Sc) du jeu de données de Cook & Johnson.

Copule sous \mathbb{H}_0	(Co,Cs)					(Co,Sc)				
	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$
N	0.234	0.248	0.468	0.398	0.930	0.723	0.612	0.714	0.862	0.874
χ^2	0.572	0.242	0.002	0.034	0.044	0.903	0.004	0.000	0.000	0.000
T_1	0.234	0.594	0.342	0.118	0.002	0.723	0.752	0.906	0.214	0.506
T_5	0.234	0.090	0.820	0.090	0.298	0.723	0.316	0.664	0.412	0.398
F_1	0.000	0.000	0.018	0.000	0.000	0.791	0.002	0.000	0.056	0.000
F_5	0.440	0.022	0.002	0.034	0.006	0.894	0.004	0.006	0.030	0.000
$P\ell$	0.326	0.172	0.450	0.110	0.072	0.851	0.876	0.140	0.210	0.086
$P\ell^2$	0.681	0.032	0.000	0.120	0.018	0.954	0.050	0.078	0.046	0.022
$C\ell$	0.354	0.048	0.338	0.184	0.856	2.120	0.000	0.000	0.000	0.000
$C\ell^2$	1.163	0.002	0.000	0.106	0.022	5.544	0.028	0.090	0.038	0.016
Fr	1.390	0.080	0.220	0.304	0.874	6.001	0.058	0.784	0.128	0.046
Fr^2	4.351	0.244	0.002	0.038	0.138	12.975	0.458	0.324	0.020	0.004
Gu	0.150	0.038	0.048	0.006	0.068	0.515	0.008	0.000	0.014	0.592
Gu^2	0.344	0.024	0.010	0.036	0.342	0.713	0.000	0.000	0.002	0.000

TABLE B.3 – P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Co,Ti) et (Sc,Ti) du jeu de données de Cook & Johnson.

Copule sous H_0	(Co, Ti)					(Sc, Ti)				
	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$
N	0.528	0.722	0.826	0.712	0.396	0.621	0.060	0.748	0.120	0.396
χ^2	0.804	0.008	0.000	0.142	0.072	0.855	0.254	0.024	0.030	0.028
T_1	0.528	0.444	0.420	0.044	0.002	0.621	0.326	0.894	0.030	0.220
T_5	0.528	0.750	0.916	0.480	0.702	0.621	0.204	0.516	0.320	0.414
F_1	0.311	0.016	0.008	0.084	0.000	0.619	0.234	0.004	0.016	0.002
F_5	0.775	0.030	0.000	0.028	0.074	0.837	0.038	0.000	0.078	0.048
$P\ell$	0.678	0.424	0.426	0.932	0.040	0.767	0.058	0.170	0.130	0.028
$P\ell^2$	0.888	0.098	0.002	0.118	0.168	0.924	0.356	0.146	0.112	0.128
$C\ell$	1.095	0.020	0.000	0.002	0.010	1.487	0.020	0.000	0.022	0.044
$C\ell^2$	3.107	0.068	0.000	0.288	0.282	4.071	0.224	0.140	0.078	0.172
Fr	3.555	0.854	0.078	0.188	0.734	4.539	0.034	0.480	0.136	0.172
Fr^2	8.355	0.424	0.002	0.008	0.644	10.188	0.584	0.004	0.004	0.042
Gu	0.354	0.330	0.394	0.174	0.002	0.427	0.012	0.010	0.266	0.802
Gu^2	0.581	0.026	0.000	0.164	0.004	0.644	0.082	0.000	0.022	0.440

TABLE B.4 – P-valeurs des tests basés sur $S_{n,N}^{\lambda=3}$, $T_{n,N}^{\lambda=3}$, $V_{n,N}^K$ et $V_{n,N}^C$ pour les paires de variables (Cs,Sc) et (Cs,Ti) du jeu de données de Cook & Johnson.

Copule sous \mathbb{H}_0	(Cs,Sc)					(Cs,Ti)				
	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$	$\hat{\theta}_n$	$S_{n,N}^{\lambda=3}$	$T_{n,N}^{\lambda=3}$	$V_{n,N}^K$	$V_{n,N}^C$
N	0.336	0.586	0.400	0.210	0.016	0.450	0.002	0.036	0.206	0.002
χ^2	0.670	0.014	0.000	0.012	0.008	0.755	0.770	0.478	0.836	0.242
T_1	0.336	0.026	0.104	0.212	0.234	0.450	0.208	0.022	0.096	0.156
T_5	0.336	0.192	0.162	0.252	0.468	0.450	0.472	0.000	0.364	0.082
F_1	0.000	0.000	0.000	0.000	0.002	0.000	0.512	0.214	0.064	0.046
F_5	0.594	0.010	0.000	0.022	0.020	0.711	0.266	0.430	0.938	0.424
$P\ell$	0.460	0.068	0.700	0.004	0.796	0.595	0.304	0.000	0.548	0.238
$P\ell^2$	0.776	0.040	0.008	0.000	0.108	0.850	0.128	0.418	0.672	0.448
$C\ell$	0.558	0.554	0.652	0.004	0.226	0.845	0.000	0.000	0.000	0.000
$C\ell^2$	1.722	0.040	0.000	0.002	0.038	2.484	0.986	0.370	0.570	0.360
Fr	2.047	0.962	0.036	0.002	0.248	2.900	0.088	0.018	0.104	0.062
Fr^2	5.594	0.368	0.250	0.008	0.064	7.143	0.038	0.020	0.248	0.242
Gu	0.218	0.150	0.002	0.006	0.138	0.297	0.776	0.062	0.140	0.072
Gu^2	0.437	0.008	0.000	0.000	0.000	0.525	0.764	0.902	0.258	0.782